

ARCHITECTURE

A write-path architecture for an AI-native EMR

ABSTRACT

An electronic medical record is a legal document. The decisive design choice is therefore not the model but the write path: what may enter the record, in what state, on whose authority, and with what provenance. This paper specifies one such path. It treats author, state and authority as separable properties of every entry, holds the AI draft as a durable object anchored to the recorded transcript, and enumerates what the path must refuse. The architecture makes generated text attributable and reviewable; it does not make it accurate.

Type Original technical description **References** 28 **Reading time** 14 min

Last reviewed September 2026

1 The write path, not the model

Most products described as AI-native electronic medical records are a model attached to a record system designed on the assumption that a human typed everything. Its schema, permissions, audit log and notion of authorship all predate the new author. The generative component is given somewhere to put its output, usually a text field that already existed, and engineering attention goes to prompts and evaluation.

The research literature has the same emphasis. A systematic review of 519 studies of large language models in health care found that only 5% used real patient care data while 44.5% evaluated on medical examination questions, and that the administrative work record systems actually mediate was close to absent, at 0.2% of studies each for billing codes and prescriptions.¹ That is a literature about model capability, not records.

This paper takes the opposite order. In an EMR the record is the product: a legal instrument, the substrate of a claim, evidence in a liability case, and what the next clinician relies on. The first question is not what a model can generate. It is where a model may write, what state its output occupies on arrival, who may move it out of that state, and what has to be true of the record before anything is committed. That set of rules is the write path.

This paper describes one write path, as implemented in ChartVoyant, with the reasoning behind each constraint. The evidence comes from the published literature, not from measurements of this system: no performance claim about ChartVoyant appears here, because none has been published.

2 Author, state, and authority

Every entry carries three properties. They are independent, and most failure modes in AI documentation collapse two of them into one.

Author

The author is the named agent that produced the content, person or machine, recorded on the entry rather than inferred from the field the text landed in. This is the *Agent* of the W3C provenance model, distinct from the *Activity* that produced the *Entity* — described here from PROV-Overview, a Working Group Note; the normative document is PROV-DM.² FHIR splits the work: Provenance covers generation, AuditEvent covers usage, so recording authorship in one and expecting it to answer access questions leaves a gap. Provenance is Trial Use, not normative.³

State

State is the entry's standing in the legal record. Two are sufficient: *draft*, recorded and durable but not part of the record a clinician is accountable for, and *attested*, meaning a clinician has taken responsibility for it. State belongs to the entry, not to whoever wrote it: human-typed text is unattested until signed, exactly as machine-drafted text is.

Authority

Authority is the right to move an entry between states. It belongs to a person, scoped to a role and a practice, never to an author by virtue of authorship. FDA's electronic records rule, which governs FDA-regulated records rather than ambulatory EMRs, states most clearly what a signature must carry: the signer's printed name, the date and time, and the *meaning* of the signature, whether review, approval, responsibility or authorship.⁴ It is an unflattering comparator.

The applicable American rule is thinner. Medicare's signature guidance states that a scribe documenting entries on a provider's behalf need not sign or date the documentation; the treating clinician's signature affirms that the note adequately documents the care provided.⁵ Applied to a machine drafter, that puts the whole attestation burden on the signer and requires no disclosure of what produced the text.

Why collapsing them fails

Collapse author into authority and the component that writes becomes the component that decides. Collapse state into author and the record acquires a false rule, that machine output is provisional and human output final, which breaks the moment a clinician edits a draft. Collapse author into state and the record forgets who wrote a sentence at the moment it is signed, the most damaging of the three: attestation is exactly when provenance becomes worth having.

None of the three is supplied by regulation. HIPAA's audit controls standard is one sentence requiring mechanisms that record and examine activity in systems holding electronic protected health information, with no content list, no retention period and no review cadence, and the mechanism for corroborating that information has not been improperly altered is Addressable, the weakest tier.⁶ Reviewing activity records is Required, but "regularly" is undefined.⁷ What specificity exists comes from certification⁸ and its audit-log content standard.⁹ A proposed Security Rule overhaul was still a proposal as of September 2026.¹⁰

3 Draft as a first-class state

It follows that the model's output is a durable, recorded object that is not part of the legal record until a clinician acts on it. In ChartVoyant the transcript fills during the visit and a draft is produced alongside it. Nothing signs itself: the draft stays a draft until a clinician reviews, signs and locks it into the record.

Three event types carry the visit across that boundary. `visit.transcript.appended` records speech captured during the visit. `ai.note.drafted` records the machine-written draft and links to the transcript it came from. `note.signed` records the signing clinician and the timestamp. Every action, including the model's, goes to a log that can only be added to, never edited or deleted, and each entry carries its own digest and a pointer to the previous entry's digest, so altering history is detectable. Records, files and audit logs are separated per practice; applied here, that yields a per-practice chain.

The construction is old and needs no distributed ledger: Haber and Stornetta's linking scheme chains documents through cryptographic hashes, making back-dating computationally infeasible.¹¹ Chaining seals entries against one another, but the times in the chain are the system's own; an external time reference is a separate guarantee, obtained from a timestamping authority that signs over a hash without seeing the data behind it.¹²

Append-only is not immutable, and the difference matters clinically: records must stay correctable, since clinicians retract, addenda are added and patients may request amendment. What is needed is append-only with tamper-evidence.

Why the draft is written to the log

Holding a draft in memory until a clinician accepts or discards it is simpler, cheaper and wrong. An unreviewed suggestion that disappears cannot be audited: if the only drafts that survive are the ones a clinician approved, the log records the system's successes and nothing else. A rejected draft is evidence too.

The reviewer's decision also means nothing if the object decided upon was not preserved. An attestation is a statement about a specific artifact, and a signature over a draft that no longer exists, or that the reviewer mutated in place, is a signature over nothing recoverable. Keeping the draft as one entry and

the signature as another is what allows the difference between them to be reconstructed, which is the only way to establish what the clinician changed.

4 Why the transcript is the anchor

The link from `ai.note.drafted` back to `visit.transcript.appended` is the load-bearing part of the design. Review is only a control if verification is cheap.

Lyell and Coiera analyzed 40 studies from nine databases covering 1983 to 2015, only 6 of them healthcare-focused. Automation bias was not principally a multitasking artifact: it appeared in single tasks, typically diagnosis rather than monitoring, and where verification complexity was high.¹³ That names a variable the designer controls.

Leaving verification expensive has a measured consequence. In a randomized experiment with 120 final-year medical students, incorrect decision support raised total prescribing errors by 86.6% relative to no support at all, while correct support reduced them by 58.8%.¹⁴ In a single-blind randomized trial of 44 physicians who had all received AI-literacy training, deliberately flawed language-model output cut diagnostic reasoning scores by an adjusted -14.0 percentage points (95% CI -19.7 to -8.3 ; $P < .0001$).¹⁵ Training did not protect them. Putting the source document one click from the generated sentence acts on the variable the evidence identifies.

The failures worth designing against are counted. Across 450 clinical note and transcript pairs, clinicians annotated 12,999 sentences for hallucination and 49,590 for omission: 1.47% of sentences were hallucinated, 44% of those graded major, and 3.45% were omissions. Major hallucinations clustered in the Plan section, at 21%.¹⁶ The highest-risk fabrications sit in the part of the note that says what happens next.

TABLE 1

Reported error rates in generated or transcribed clinical text use incompatible denominators. A write path must be specified against the unit that matters to it: the sentence a clinician will attest to.

Study	Unit of measurement	Reported rate
Asgari et al. 2025 ¹⁶	Sentences of a generated note	1.47% hallucinated (44% major); 3.45% omissions
Palm et al. 2025 ¹⁷	Notes containing at least one hallucination	31% ambient vs 20% physician-drafted
Omar et al. 2025 ¹⁸	Vignettes seeded with one fabricated element	66% overall; 50–82.7% by model
Koenecke et al. 2024 ¹⁹	Transcriptions of non-medical speech corpora	~1% held a fabricated phrase; 38% of those harmful
Zhou et al. 2018 ²⁰	Words of dictated clinical documents	7.4 errors per 100 words at recognition output

Grounding does not make a draft correct. In an adversarial study of 300 physician-validated vignettes, each seeded with one fabricated element, six models elaborated on the fabrication rather than rejecting it, 66% of the time overall; a mitigation prompt cut that to 44%, and temperature zero produced no significant improvement.¹⁸ A model handed a false premise builds on it.

The anchor is itself a recording rather than ground truth. An audit of a widely used speech-to-text system found entirely hallucinated phrases in roughly 1% of transcriptions, 38% of them containing explicit harms, disproportionately triggered by longer non-vocal pauses.¹⁹ Those corpora were aphasia and control speech, not medical dictation, so that is not a medical-setting error rate. The design rests on something narrower: the transcript is the recorded input the draft came from, and what a reviewer checks a sentence against.

5 What the write path must refuse

The specification is easier to state as prohibitions; four refusals define it.

1. No silent writes. Every entry names its author, and a machine author is named as such rather than inheriting the identity of the session or the clinician it ran for. Certification requires actions on electronic health information to be recorded against an adopted content standard,⁹ but nothing requires a machine-drafted narrative to be identified as machine-drafted: the source-attribute disclosures reach decision support interventions and stop there.⁸ A system closes that gap by making the drafting event a first-class event type with its own author, which is what `ai.note.drafted` is.

2. No model-initiated state transitions. Only a licensed clinician may move an entry from draft to attested, including entries the model authored. Clinical suggestions are drafts; the model decides nothing on its own. A wrong suggestion a reviewer accepts is worse than no suggestion,^{14,15} so no path to the record may bypass a reviewer. Because the signing clinician's attestation is what makes a note authoritative,⁵ any mechanism producing an attested entry without a signer manufactures authority no one granted.

3. No post-hoc edits that leave no trace. Corrections are new entries and nothing is overwritten in place. The record already has a provenance problem predating generative systems: across a decade at one academic medical center, median outpatient note length rose 60.1% and notes written in 2018 averaged 29.4% directly typed text.²¹ Text that arrived by copy or template carries no record of its origin, and adding a generative author to a record that cannot say where its sentences came from turns a quality problem into a categorical one.

4. No reconstruction of an encounter from a downstream artifact. The claim, the referral letter and the printed chart are outputs of the record; none is the encounter. Printed renderings differ in formatting and chronology from the record beneath them, which is why recreating an accurate timeline in a liability case is specialist work.²² Permitting downstream reconstruction guarantees two versions of the encounter and no way to say which is authoritative.

TABLE 2

The five publicly documented event types in ChartVoyant's audit chain. The state column is the specification described in this paper, not a product label; each entry carries its own digest and a pointer to the previous entry's digest.

Event	Author	State on write	Who may advance it
<code>visit.transcript.appended</code>	Capture of the visit	Recorded, not asserted	No one; appended to, never rewritten
<code>ai.note.drafted</code>	The model, named, linked to the transcript	Draft	The treating clinician, by signing
<code>note.signed</code>	The signing clinician, with timestamp	Attested	No one; a correction is a new entry
<code>claim.submitted</code>	Billing, projected from the attested note	Derived	The payer; its response is a new entry
<code>payment.posted</code>	Posting, from the payer's remittance	Derived	No one; adjustments are new entries

6 Downstream artifacts are projections

The last two rows of Table 2 follow from the first three. `claim.submitted` records an 837P sent to a clearinghouse and `payment.posted` records an insurer payment against it, both in the same append-only chain as the events they descend from. A claim, an order or a letter is a projection of an attested encounter, not a reconstruction of it.

Two consequences follow. The design requires that a claim be projectable only from an attested encounter, because otherwise the projection has no source. And the code set that reaches a claim is the set a clinician confirmed: ChartVoyant's drafted output includes suggested codes, illustrated publicly with ICD-10-CM M54.16 and G89.29, and those are part of the draft, subject to the same confirmation as the narrative. Billing codes were the subject of 0.2% of studies in the review cited above.¹ That is a reason to be conservative here, not to treat coding as easy.

7 What this architecture costs

The costs are real, and several are permanent rather than implementation debt.

Storage and event volume. The design requires every draft to be retained, including rejected ones, alongside the transcript it came from. The log grows with activity, not with the size of the chart, and it never shrinks.

Latency between speech and a usable note. The draft appears during the visit; the note does not exist until someone signs it. No model improvement removes that interval, because the interval is the control.

A review burden that does not go away. A randomized study of AI-generated draft replies to patient messages found read time rose 21.8% (95% CI 5.2% to 41.0%; $P=.008$) with no significant change in reply time (-5.9% ; $P=.33$).²³ A meta-analysis of human and language-model collaboration found a composite improvement of 4.88 percentage points (95% CI 0.65 to 9.12) with a prediction interval crossing null, factual error rates persisting at 26% to 36%, and collaboration not universally outperforming the model alone; the authors call the supervision cost a “vigilance tax.”²⁴ That tax is what this write path charges.

A benefit that is not categorical. A pragmatic randomized trial of 238 outpatient physicians across 14 specialties found one ambient product reduced time-in-note by 9.5% (95% CI -17.2% to -1.8% ; $P=0.02$) while a second showed no significant change (-1.7% ; $P=0.66$).²⁵ Architecture does not determine whether a product saves time. Review time is not free either: each additional hour a primary care physician spends on documentation is associated with a 7.1% decrease in the likelihood of accessing outside patient records.²⁶

A constraint on what can be automated. Auto-signing a normal note and auto-submitting a clean claim are unavailable by construction. That is the intent, and still a cost. The gap the review step exists to catch is measurable: across 5 standardized primary care cases rated by 30 blinded raters, notes from 11 AI scribe tools scored lower than notes from 18 human note takers on every case and on all 10 modified PDQI-9 domains.²⁷ Those cases were simulated, so the gap is an upper bound.

LIMITATIONS OF THIS DESIGN

The write path buys attribution and reviewability at a price paid in storage, event volume, latency and clinician attention, and it forecloses automations a less constrained system could ship. None of those costs is recovered by a better model, because none is caused by the model.

8 What it does not solve

This architecture does not make the model accurate. It makes the model’s output attributable and reviewable. Those are different properties, and the second does not imply the first.

The sharpest evidence for the distinction predates language models. Across 217 dictated clinical documents, the error rate per 100 words fell from 7.4% in raw speech-recognition output to 0.4% after transcriptionist editing and 0.3% in physician-signed notes. The proportion of remaining errors that were clinically significant did not fall: 5.7%, then 8.9%, then 6.4%.²⁰ Review removed errors; it did not preferentially remove the dangerous ones. A write path routing generated text past a reviewer inherits that result exactly.

The log has limits of its own. Across 85 studies using EHR audit logs to study clinical activity, only 19 (22%) validated their results and 9 (11%) validated against direct observation.²⁸ An audit chain is strong evidence of what the system recorded, not of what happened in the room. It makes a record's history checkable, not its contents true.

NO MEASUREMENT IS REPORTED HERE

This paper reports no accuracy, time-saving, documentation-burden, coding or denial-rate measurement for ChartVoyant, because none has been published. What is described above is a specification, not a demonstrated outcome. ChartVoyant's own security documentation states that no system is perfectly secure and that the company does not claim otherwise; the same register applies here.

9 Conclusion

The model in any AI-native record system will be replaced, probably within a year. The record will not. That asymmetry is the argument for specifying the write path first and treating the model as one more author subject to it: named, producing output in a state it cannot leave on its own, anchored to the input it came from, in a log that can be appended to and not rewritten.

What such a design yields is a set of answerable questions. What did the model produce. What was it given. Who reviewed it, and what did they change. What was sent downstream, and from which attested entry. A system that cannot answer those does not solve them by choosing a better model; one that can has not thereby made the model right.

REFERENCES

Entries 2, 3 and 12 are standards specifications, and entry 3 is Trial Use rather than normative. Entries 4 to 10 are regulations or sub-regulatory guidance rather than research; entry 10 is a proposed rule that had not been finalized as of September 2026 and is cited only as a proposal. The remainder are peer-

reviewed studies and reviews. Reported error rates in entries 16 to 20 use incompatible denominators, so the unit of measurement is given wherever a rate appears.

1. Bedi S, Liu Y, Orr-Ewing L, et al. *Testing and Evaluation of Health Care Applications of Large Language Models: A Systematic Review*. JAMA. 2025;333(4):319–328. [doi:10.1001/jama.2024.21700](https://doi.org/10.1001/jama.2024.21700) SYSTEMATIC REVIEW
2. Groth P, Moreau L, eds. *PROV-Overview: An Overview of the PROV Family of Documents*. W3C Working Group Note, 30 April 2013. [w3.org/TR/prov-overview](https://www.w3.org/TR/prov-overview) STANDARD
3. Health Level Seven International. *Provenance*. HL7 FHIR Release 5; Maturity Level 4, Trial Use. hl7.org/fhir/R5/provenance.html STANDARD
4. U.S. Food and Drug Administration. *21 CFR Part 11 — Electronic Records; Electronic Signatures*. Effective 20 March 1997. [ecfr.gov](https://www.ecfr.gov) REGULATION
5. Centers for Medicare & Medicaid Services. *Medicare Program Integrity Manual, Chapter 3, § 3.3.2.4 — Signature Requirements*. [cms.gov](https://www.cms.gov) GUIDANCE
6. U.S. Department of Health and Human Services, Office for Civil Rights. *45 CFR § 164.312 — Technical safeguards*. [ecfr.gov](https://www.ecfr.gov) REGULATION
7. U.S. Department of Health and Human Services, Office for Civil Rights. *45 CFR § 164.308(a)(1) — Security management process*. [ecfr.gov](https://www.ecfr.gov) REGULATION
8. U.S. Department of Health and Human Services, ASTP/ONC. *45 CFR § 170.315 — 2015 Edition health IT certification criteria*. [ecfr.gov](https://www.ecfr.gov) REGULATION
9. U.S. Department of Health and Human Services, ASTP/ONC. *45 CFR § 170.210(e) — Record actions related to electronic health information, audit log status, and encryption of end-user devices; incorporating ASTM E2147-18, Standard Specification for Audit and Disclosure Logs for Use in Health Information Systems*. [ecfr.gov](https://www.ecfr.gov) REGULATION
10. U.S. Department of Health and Human Services, Office for Civil Rights. *HIPAA Security Rule To Strengthen the Cybersecurity of Electronic Protected Health Information*. Notice of proposed rulemaking. 90 FR 898, 6 January 2025. [federalregister.gov](https://www.federalregister.gov) REGULATION
11. Haber S, Stornetta WS. *How to time-stamp a digital document*. Journal of Cryptology. 1991;3(2):99–111. [doi:10.1007/BF00196791](https://doi.org/10.1007/BF00196791) PEER-REVIEWED
12. Adams C, Cain P, Pinkas D, et al. *Internet X.509 Public Key Infrastructure Time-Stamp Protocol (TSP)*. IETF RFC 3161, Standards Track. August 2001. [rfc-editor.org](https://www.rfc-editor.org) STANDARD
13. Lyell D, Coiera E. *Automation bias and verification complexity: a systematic review*. Journal of the American Medical Informatics Association. 2017;24(2):423–431. [doi:10.1093/jamia/ocw105](https://doi.org/10.1093/jamia/ocw105) SYSTEMATIC REVIEW

14. Lyell D, Magrabi F, Raban MZ, et al. *Automation bias in electronic prescribing*. BMC Medical Informatics and Decision Making. 2017;17:28. doi:10.1186/s12911-017-0425-5
CONTROLLED EXPERIMENT
15. Qazi IA, Ali A, Khawaja AU, et al. *Automation Bias in Large Language Model–Assisted Diagnostic Reasoning among Physicians Trained in AI Literacy — A Randomized Clinical Trial*. NEJM AI. doi:10.1056/Aloa2501001. Preprint posted 26 August 2025: doi:10.1101/2025.08.23.25334280
RCT
16. 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 Digital Medicine. 2025;8:274. doi:10.1038/s41746-025-01670-7 BENCHMARK
17. Palm E, Manikantan A, Mahal H, et al. *Assessing the quality of AI-generated clinical notes: validated evaluation of a large language model ambient scribe*. Frontiers in Artificial Intelligence. 2025;8:1691499. doi:10.3389/frai.2025.1691499 COMPARATIVE STUDY
18. Omar M, Sorin V, Collins JD, et al. *Multi-model assurance analysis showing large language models are highly vulnerable to adversarial hallucination attacks during clinical decision support*. Communications Medicine. 2025;5:330. doi:10.1038/s43856-025-01021-3 BENCHMARK
19. Koenecke A, Choi ASG, Mei KX, et al. *Careless Whisper: Speech-to-Text Hallucination Harms*. Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency (FAccT '24). arXiv:2402.08021. arxiv.org/abs/2402.08021 BENCHMARK
20. Zhou L, Blackley SV, Kowalski L, et al. *Analysis of Errors in Dictated Clinical Documents Assisted by Speech Recognition Software and Professional Transcriptionists*. JAMA Network Open. 2018;1(3):e180530. doi:10.1001/jamanetworkopen.2018.0530 CROSS-SECTIONAL
21. Rule A, Bedrick S, Chiang MF, et al. *Length and Redundancy of Outpatient Progress Notes Across a Decade at an Academic Medical Center*. JAMA Network Open. 2021;4(7):e2115334. doi:10.1001/jamanetworkopen.2021.15334 LONGITUDINAL ANALYSIS
22. Sittig DF, Wright A. *Identifying a Clinical Informatics or Electronic Health Record Expert Witness for Medical Professional Liability Cases*. Applied Clinical Informatics. 2023;14(2):290. doi:10.1055/a-2018-9932 POSITION PAPER
23. Tai-Seale M, Baxter SL, Vaida F, et al. *AI-Generated Draft Replies Integrated Into Health Records and Physicians' Electronic Communication*. JAMA Network Open. 2024;7(4):e246565. doi:10.1001/jamanetworkopen.2024.6565 RANDOMIZED QI STUDY
24. Wang G, Zhang K, Jiang J, et al. *Human–large language model collaboration in clinical medicine: a systematic review and meta-analysis*. npj Digital Medicine. 2026;9:195. doi:10.1038/s41746-026-02382-2 META-ANALYSIS

25. Lukac PJ, Turner W, Vangala S, et al. *Ambient AI Scribes in Clinical Practice: A Randomized Trial*. NEJM AI. 2025;2(12). doi:10.1056/Aloa2501000 RCT
26. Holmgren AJ, Adler-Milstein J, Apathy NC. *Electronic Health Record Documentation Burden Crowds Out Health Information Exchange Use By Primary Care Physicians*. Health Affairs. 2024;43(11):1538–1545. doi:10.1377/hlthaff.2024.00398 COHORT
27. Reddy A, Gunnink E, Wheat CL, et al. *Rapid Evaluation of Artificial Intelligence Technology Used for Ambient Dictation in Primary Care: Comparing the Quality of Documentation of Artificial Intelligence-Generated and Human-Produced Clinical Notes*. Annals of Internal Medicine. 2026;179(6):765–772. doi:10.7326/ANNALS-25-02772 CROSS-SECTIONAL
28. Rule A, Chiang MF, Hribar MR. *Using electronic health record audit logs to study clinical activity: a systematic review of aims, measures, and methods*. Journal of the American Medical Informatics Association. 2020;27(3):480–490. doi:10.1093/jamia/ocz196 SYSTEMATIC REVIEW

RELATED PAPERS

Approval gates

The state transition where authority actually changes hands.

Tamper-evident clinical records

The hash chain that makes the write path's history checkable.

Retrieval over the chart

The read path, and the data quality it has to survive.