Internet Medical Journal
ISSN 1093-7935 · medjournal.com · Est. 1996

Blockchain Audit Trails Resolve the Electronic Health Record Traceability Problem Created by Generative AI

Author(s)	Thomas F. Heston
Affiliation(s)	Department of Family Medicine, University of Washington, Seattle, USA
Affiliation(s)	Department of Medical Education and Clinical Sciences, Elson S. Floyd College of Medicine, Washington State University, Spokane, USA
ORCID	0000-0002-5655-2512

Published	21 May 2026
DOI	10.5281/zenodo.20337356
Article type	Commentary
Citation	Heston TF. Blockchain Audit Trails Resolve the Electronic Health Record Traceability Problem Created by Generative AI. Internet Medical Journal. 2026;1:e20337356

Abstract

As large language models generate clinical documentation at scale, electronic health records increasingly contain AI-produced content with no verifiable provenance. The field has named this traceability gap but has not yet specified an architectural solution. Blockchain-based audit logging — append-only, cryptographically chained, and tamper-resistant — provides the answer: a lightweight layer capturing model identity, prompt hash, output fingerprint, and timestamp at generation, creating a verifiable chain of custody before text enters the clinical record. Adopting blockchain audit logging as a standard condition of institutional large language model deployment would resolve this traceability crisis before it becomes irreversible.

Keywords

electronic health records, generative AI, blockchain, audit trail, traceability, large language models, clinical governance, provenance

Large language models are generating clinical notes, orders, and summaries that are entered into electronic health records (EHRs) without a verifiable record of their origin, and blockchain-based audit logging provides the architectural solution the field requires. The traceability problem is not incidental to large language model deployment in clinical settings — it is intrinsic to how these systems operate. A model processes a prompt and returns output; that output may be edited, accepted wholesale, or lightly paraphrased before entering the record, and current EHR audit functions capture none of this chain of events. The clinical and regulatory consequences of this gap are not speculative: when a physician's note contains model-generated content, questions of authorship, accountability, and revision history become legally and clinically material.

Blockchain technology was designed precisely for this class of problem. Its defining property is the append-only, cryptographically chained distributed ledger: each record is linked to all prior records by a cryptographic hash, making retroactive alteration computationally infeasible without detection. This property has been validated across healthcare settings — in health information exchange, federated learning, pharmaceutical supply chain verification, and national eHealth infrastructure [1]. Blockchain does not require storing clinical content on a public ledger — permissioned architectures maintain institutional privacy while preserving the cryptographic guarantees that make audit trails meaningful. The technology is not experimental; it is deployable at scale today. Given that blockchain already possesses precisely the properties this problem requires, the absence of a named architectural solution in the clinical literature represents an unmet match — not a gap in technology, but a gap in application.

Existing EHR audit mechanisms are insufficient for the large-language-model era [2]. Current audit logs capture user identity, timestamp, and edit history — they record who touched the record and when. What they cannot capture is whether a note was drafted by a model, which model version produced the output, what prompt was submitted, or how substantially the output was modified before being accepted. This has been described as a "blending" problem: AI- and human-generated content become indistinguishable within the clinical record, with no mechanism to recover the distinction after the fact. The published literature calls explicitly for "technological and policy solutions to ensure traceability of AI-generated content in EHRs to preserve clinical integrity" — but stops short of specifying an architecture. The gap is therefore not diagnostic but prescriptive: the problem has been named; the solution has not.

Given that blockchain's core properties — append-only ledger, cryptographic chaining, tamper-resistant distributed storage — have been validated for clinical data provenance, it follows that these same properties resolve this traceability gap. A blockchain-based audit layer does so by logging four fields at the moment of model output generation, before text

enters any clinical workflow: (1) model identifier and version, (2) a cryptographic hash of the submitted prompt, (3) a cryptographic hash of the unedited model output, and (4) a precise timestamp. This four-field construct constitutes Heston's blockchain audit standard for large language model governance — a minimal, privacy-preserving architecture deployable within existing institutional infrastructure. These four fields, recorded to a permissioned blockchain ledger, create a verifiable chain of custody. The chain holds regardless of whether, how, or how much the clinician subsequently edited the output. The hash of the original output does not prevent editing. It preserves the ability to verify, on audit, whether the final note matches the model's original generation. This architecture does not require storing clinical content on the blockchain, only its cryptographic fingerprint, making it compatible with existing HIPAA-compliant infrastructure. A cryptographic hash, such as SHA-256 (the algorithm underlying Bitcoin), is a one-way mathematical transformation: it converts any text into a fixed-length string — for example, a3f8c2... — from which the original content cannot be reconstructed. No patient name, diagnosis, or clinical text resides on the ledger, nor can they be reconstructed from the ledger; privacy is 100% preserved. Verification works by re-running the stored note through the same algorithm and comparing outputs. A match confirms the note is unaltered. A mismatch flags a discrepancy. In either case, the clinical content itself never leaves the institution's secure environment. One objection warrants a direct answer. Existing EHR audit logs or centralized server logs could theoretically be extended to capture large-language-model provenance. This argument fails on tamper-resistance. Centralized logs can be altered by administrators, overwritten during system migrations, or lost to infrastructure failures — single points of failure that blockchain architecture was designed to eliminate. Cryptographic chaining ensures that any retroactive modification of a logged event is detectable without requiring trust in any single institution. The governance argument for this approach has been developed in detail elsewhere [3].

Three lines of evidence support the feasibility of this approach. First, national-scale deployment of blockchain-secured health records is not hypothetical: Estonia has operated KSI (Keyless Signature Infrastructure) blockchain-secured eHealth infrastructure since 2011 in partnership with Guardtime, making it one of the longest-running real-world demonstrations of tamper-resistant clinical data integrity at a population level [4,5]. Second, a quantum-secure healthcare blockchain architecture has been demonstrated to support medical data logging at 105 transactions per second with low CPU overhead and linear scalability under simulated load [6], well within the throughput requirements of institutional large-language-model deployments. Third, the urgency of this solution is established by the pace of institutional adoption: a privately hosted large language model deployed at a radiology institution achieved 93.1% protocol accuracy on MRI examination requests, matching board-certified radiologists, with outputs entering the clinical workflow in real time [6]. At that scale and accuracy, the absence of an audit trail is not a theoretical concern — it is an active governance gap. Evidence-based frameworks for integrating generative AI

with blockchain governance are available to guide institutional implementation [7]. Given this evidence of technical maturity, national-scale deployment, and accelerating clinical adoption, the barrier to implementation is not capability — it is policy.

Clinical institutions deploying large language models in EHR-connected workflows should adopt blockchain audit logging as a condition of deployment, with four fields logged at the point of generation: model identifier and version, cryptographic hash of the prompt, cryptographic hash of the unedited output, and timestamp. This standard does not constrain clinical workflow — the clinician retains full authority to edit, reject, or accept model output — but it preserves the chain of custody that clinical governance, legal accountability, and regulatory review require. Journals publishing studies of institutional large language model deployment in clinical settings should require disclosure of whether audit logging was implemented, on what infrastructure, and whether logs were reviewed as part of study quality control. Text-based AI can now reason like a physician; the remaining challenge is achieving safe clinical implementation [8]. Blockchain audit trails are a concrete, technically mature component of that implementation — and one the field is equipped to adopt now.

Declarations

Funding: This study did not receive any external funding.

Conflicts of Interest: The author reports no conflicts of interest.

Data Availability: Not applicable.

Research Ethics Statement: Not applicable. This commentary did not involve human subjects research, animal research, or protected health information.

AI Usage: Large language models were used for language editing and formatting assistance; the author reviewed, verified, and is fully responsible for all content.

References

1. Heston TF. The blockchain-based scientific study. Digit Med. 2017;3: 66. doi:10.4103/digm.digm_17_17

2. Nargesi AA, You JG, Bitterman DS, Succi MD, Mishuris RG, Poon EG, et al. Tracing the Pen: Electronic Health Records Amid the Rise of Generative AI. Npj Digit Med. 2026 [cited 21 May 2026]. doi:10.1038/s41746-026-02508-6

3. Heston TF. Accountable Clinical AI Requires More Than Accuracy. Internet Med J. 2026;1: e19519377–e19519377. doi:10.5281/zenodo.19519377

4. Soares B, Ferreira A, Veiga PM. The Benefits and Challenges of Blockchain Technology and eHealth Implementation in Estonia - A Literature Review. Appl Med Inform. 2023;45. Available: https://ami.info.umfcluj.ro/index.php/AMI/article/view/990

5. Semenzin S, Rozas D, Hassan S. Blockchain-based application at a governmental level: disruption or illusion? The case of Estonia. Policy Soc. 2022;41: 386–401. doi:10.1093/polsoc/puac014

6. Hallinan JTPD, Leow NW, Low YX, Lee A, Ong W, Chan MDZ, et al. Initial Insights Into an Institutional Secure Large Language Model for Magnetic Resonance Imaging Examination Requests: Retrospective Study. J Med Internet Res. 2026;28: e82579. doi:10.2196/82579

7. Heston TF. Evidence-Based Frameworks for Generative Artificial Intelligence. Int J Blockchain Technol Appl. 2026;4: 34–38. doi:10.5281/ZENODO.18898203

8. Hopkins AM, Cornelisse E. AI can reason like a physician—what comes next? Science. 2026;392: 466–467. doi:10.1126/science.aeg8766