BLOCKCHAIN® IN HEALTHCARE TODAY
PLATFORM APPROACHES IN HEALTHCARE

OPINIONS/PERSPECTIVES/POINT OF VIEW

# The Most Dangerous Thing in Healthcare Isn't AI. It's What's Underneath It

Jim Schwoebel[1,2] and Tricia Wang[3]

[1]Research Director, Advanced AI Society, Chair, Task Force for AI Agents in Healthcare; [2]CEO, Quome, Advanced AI Society, Los Angeles, California, USA; [3]Advanced AI Society, Brooklyn, New York, USA

DOI: https://doi.org/10.30953/bhty.v9.529

Corresponding Author: Tricia Wang, Email: tricia@advancedaisociety.org

Submitted: July 24, 2026, Accepted: July 30, 2026, Published: August 26, 2026

Hospitals and insurers face financial pressure to use artificial intelligence (AI) but are deploying it on top of older systems that cannot show what the AI actually did. Documented cases show algorithms denying care at very high error rates with almost no outside check. Autonomous AI agents make this worse, because they can act differently from one run to the next. Deterministic, explainable AI tools and compliance-ready cloud platforms already exist for regulated healthcare, and they work with the existing hospitals' infrastructure. What is still missing is agreement on what 'verified' means: a shared, open, independent standard, not a vendor's word that its own system is safe.

Key issues:

- Financial strain is pushing adoption of AI, and oversight has not kept pace.
- Documented AI-assisted denials show high reversal rates on appeal, alongside very low appeal rates.
- Non-deterministic AI agents cannot be certified once; they require continuous verification.
- Deterministic, explainable AI tooling, cryptographic methods, and decentralized solutions are already in production in regulated life sciences and interoperate with hospitals' existing cloud infrastructure.
- The remaining gap is a shared, open, independent standard that defines 'verified' because verification cannot belong to the party being verified.

Mr. Jones, a 74-year-old man, had a stroke in October 2022. His doctor recommended extended nursing home care. UnitedHealth and its subsidiary NaviHealth cut off his coverage, restored it on appeal, then cut it off again, a cycle that repeated over months.[1,2] He and his wife spent more than $70,000 out of pocket. He died in an assisted living facility about a year later.[1]

This is not a story about one bad algorithm. It is a story about opacity compounding opacity. Long before AI, healthcare's claims and authorization systems were already black boxes, decisions made by machinery that no patient, doctor, or regulator could inspect. Now the industry is racing to deploy autonomous AI on top of that same unaudited foundation, adding a second, faster kind of opacity instead of fixing the first. The most dangerous thing in healthcare right now is not that AI is deciding too much. It is that neither the old system nor the new one can show what it did.

Mr. Jones is not an outlier. UnitedHealth's nH (naviHealth) Predict algorithm carried an alleged 90% error rate on appeal, yet only about 0.2% of patients ever appealed, and its internal workings were never made public.[1,3]

UnitedHealth is not the only insurer leaning on automation at this scale: at Cigna, another top national insurer, a separate review system let medical directors reportedly deny more than 300,000 claims in 2 months, averaging 1.2 seconds of review per denial.[4] Two different insurers, two different systems, the same pattern: no one outside the company could show what either system had actually done.

Healthcare cannot opt out of the pressure behind these denials – a pressure that squeezes hospitals as hard as it squeezes patients. Median hospital operating margins have run negative into 2026, with drug and supply costs up near 8% year over year.[5] The same coverage-and-authorization machinery that cut off our patient's care also denies hospitals' own claims for payment: final claim denials rose from 2.5% to 2.7% of billed claims between 2024 and 2025, helping drive net revenue leakage from $38.6 billion to $48.4 billion.[5,6] Financial strain and the accountability gap are not two problems. They are one.

Now the industry is selling autonomous agents as the fix. That is exactly backwards: the industry is tasking opaque agents to read charts, write orders, and act across tools on their own, on top of infrastructure that was already opaque before a single agent was deployed. You cannot fix an unaudited system by adding an unaudited agent to it. The first step, whatever remains true of the systems underneath, is to verify what the agent actually did.

Cigna's own defense proves the point. The company itself has said PxDx (procedure to diagnosis) is not AI, calling it similar to software insurers have used for years, and Elizabeth

Edwards of the National Health Law Program has noted that much of the documented harm comes from simple rules-based engines that may not meet any technical definition of AI.[7] If crude automation, not even real AI by the company's own account, already produces harm no one can audit, agentic systems will multiply it. Memory-enabled agents can be corrupted by an instruction hidden inside an ordinary document weeks before it ever executes, evading defenses built to catch attacks in the moment.[8] Open-source countermeasures built specifically for clinical settings are already in testing: one preprint describes an inline firewall that catches both the injection and the out-of-scope, protected health information (PHI)-exposing request a generic detector misses, cutting a mock deployment's harmful-action rate to near zero.[9] And agents are non-deterministic, meaning the same agent, given the same inputs, can act differently from one run to the next.

The same opacity runs across healthcare's full supply chain, from the hospital bedside to the drug pipeline, and chief information security officers are already living it at scale.

> 'In large-scale AI deployments, organizations increasingly require verifiable evidence rather than vendor-reported assertions. When activity records can be altered and runtime behavior is not fully observable, trust, accountability, and auditability become significant concerns. Establishing independent visibility into system actions is therefore becoming a foundational requirement for AI governance, operational assurance, and regulatory readiness.'

Charles Iheagwara, Global Head of AI and Cybersecurity at AstraZeneca.[10]

For years, security locked the perimeter, which made it the office that slowed everything down. Agents change that arithmetic: lock one down completely, and it does nothing useful; open it up without verification and the organization is exposed. Continuous, evidence-based verification is the way out, and the same tamper-evident record that shows what an agent did is the record that later defends a denial or supports a fraud case.

Healthcare carries one constraint most industries do not face: opacity cannot be fixed by exposing the data because the data belongs to the patient, and the methods behind it belong to the company that built them. The goal is to verify that an agent actually did what it was authorized to do, without compromising privacy, which is precisely what deterministic, auditable reasoning and cryptographic, decentralized tooling are built for. Zero-knowledge methods, decentralized records, explainable models, and trusted execution environments can confirm an agent stayed inside its authorization and show what it did, without a single patient record on display.

Some will say this kind of verifiable, auditable AI infrastructure is not ready or that getting it would mean tearing out a hospital's existing systems at exorbitant cost. Neither is true, and the mix of technology already available makes both points at once: deterministic, certified tooling is already screening pharmaceutical marketing claims for compliance inside one life sciences company's platform, and, separately, embedding patient identity verification directly into electronic health record systems like Epic.[11,12] None of this asks hospitals or healthcare to rebuild their infrastructure to start implementing verifiability now.

What none of it does, on its own, is agree on what 'verified' means. Interoperability with the hospital's cloud is not the same question as whether the checking itself can be trusted, and a dozen compatible vendors each grading their own homework is still a dozen assertions, not evidence. Verification cannot belong to the party being verified. Open source lets the world see how AI was built. Open verification, an independent, inspectable standard for how the checking is done, is how we see what it did. That discipline is now being organized under the name Proof-of-Control.

The precedent already exists in law, a register this audience will recognize. The Guiding and Establishing National Innovation for U.S. Stablecoins Act (GENIUS) Act makes stablecoin issuers' claims about their reserves checkable through independent audits and public disclosure, rather than dependent on the issuer's word.[13] We already require that of the institutions holding our money. We have not yet required it of the systems deciding our health, even though healthcare carries more sensitive data and higher stakes for a person's agency than nearly any other sector.

Every hospital, insurer, and vendor is inventing its own way to check its own agents privately and certifying itself. That is the most expensive path available and not the safest one. The alternative is not to reject AI. It is to agree, in the open, on what verification means, so no single company controls the answer.

None of this brings back what Mr. Jones and his wife spent fighting a system that was opaque before AI ever touched it. It can make sure the next denial and the next agent's action do not disappear into a system that is opaque twice over.

## Funding

None.

## Financial and Non-Financial Relationships and Activities

None.

## Data Availability Statement (DAS), Data Sharing, Reproducibility, and Data Repositories

Does not apply.

## Application of AI-Generated Text or Related Technology

None disclosed.

## Contributions

The authors are responsible for all aspects of the artcle.

## References

1. Estate of Gene B. Lokken et al. v. UnitedHealth Group Inc. et al. [Internet]. Case No. 0:23-cv-03514. U.S. District Court, District of Minnesota. Complaint filed 2023 Nov 14 [cited 2026 Jul 22]. Available from: https://litigationtracker.law.georgetown.edu/litigation/estate-of-gene-b-lokken-the-et-al-v-unitedhealth-group-inc-et-al/
2. Douglas M, Anderson R. When AI decides you don't need medical care. Barking Justice Media, The Firing Line [Internet]. 2026 Feb 11 [cited 2026 Jul 22]. Available from: https://thefiringline.substack.com/p/when-ai-decides-you-dont-need-medical
3. Ross C, Herman B. UnitedHealth faces class action lawsuit over algorithmic care denials in Medicare Advantage plans [Internet]. STAT News. 2023 Nov 14 [cited 2026 Jul 22]. Available from: https://www.statnews.com/2023/11/14/unitedhealth-class-action-lawsuit-algorithm-medicare-advantage/



Citation: Blockchain in Healthcare Today 2026, **9**: 529 - https://doi.org/10.30953/bhty.v9.529

4. ProPublica. How Cigna saves millions by having its doctors reject claims without reading them [Internet]. [cited 2026 Jul 22]. Available from: https://www.propublica.org/article/cigna-pxdx-medical-health-insurance-rejection-claims
5. Healthcare Financial Management Association (HFMA). Hospital margins decline in 2026 as expenses outpace revenue [Internet]. [cited 2026 Jul 22]. Available from: https://www.hfma.org/operations-management/hospital-operating-margin-trends-2026/
6. Healthcare Finance News. Hospitals' net revenue leakage increases 25% due to denied claims [Internet]. [cited 2026 Jul 22]. Available from: https://www.healthcarefinancenews.com/news/hospitals-net-revenue-leakage-increases-25-due-denied-claims
7. Bloomberg Law. AI, algorithm-based health insurer Denials pose new legal threat [Internet]. [cited 2026 Jul 22]. Available from: https://news.bloomberglaw.com/daily-labor-report/ai-algorithm-based-health-insurer-denials-pose-new-legal-threat
8. Dong Y, Xu S, He P, Li Y, Tang J, Liu T, et al. Memory injection attacks on LLM agents via query-only interaction [Internet]. arXiv. 2025 [cited 2026 Jul 22]. Available from: https://arxiv.org/abs/2503.03704
9. Schwoebel J, Semenec I, Rousseva J, Frasch MG, Thorstenson R, Bhatt M. Beyond injection detection: a positive-security prompt firewall that closes the scope and PHI gap SOTA classifiers miss in healthcare. medRxiv. 2026 Jun 11 [preprint, not peer reviewed]. https://doi.org/10.64898/2026.06.04.26354950
10. Iheagwara C. Personal communication, July 15, 2026.
11. Rainbird Technologies. Klick guardrail case study [Internet]. 2026. [cited 2026 Jul 22]. Available from: https://guardrail.klick.com
12. Vouched. HHS and epic [Internet]. [cited 2026 Jul 22]. Available from: https://www.vouched.id/hhs-epic
13. Krause D, Krause E. Auditing payment stablecoins under the GENIUS act [Internet]. The Regulatory Review. 2025 [cited 2026 Jul 22]. Available from: https://www.theregreview.org/2025/11/17/krause-krause-auditing-payment-stablecoins-under-the-genius-act/