AMERICAN HEALTH LAW ASSOCIATION

July 11, 2025
Health Law Weekly

Blame It on the Bot: Health Care Fraud and Compliance in the Age of AI

July 11, 2025

Joshua Robbins, Buchalter | **Daniel Pietragallo**, Buchalter

A large health system implements an Artificial Intelligence (AI)-powered clinical decision support tool that promises to ensure "complete and accurate" diagnosis documentation. Within six months, the system's Medicare reimbursements have increased by 40%. The reason: the AI has learned to identify and suggest every possible

secondary diagnosis that increases reimbursement—transforming routine elderly care into complex cases requiring maximum payment rates. The AI taught itself to game the billing system.

A medical device maker supplies an AI platform to surgery practices, designed to help support treatment and recovery plans. The device company also pays the practices a regular fee for entering follow-up data to help train and improve the AI model. The AI consistently recommends the device company's products for a large percentage of each practice's cases.

These scenarios could arise from unanticipated "emergent" behavior by AI systems that are, by nature, not fully predictable.[1] Or they could be examples of "intent laundering"—using AI to wash away criminal intent from health care fraud schemes. Unlike traditional fraud, which requires explicit agreements and clear intent, AI-powered schemes can achieve the same illegal outcomes while providing plausible deniability to all parties involved. The defense: the algorithm did it.

This emerging threat is forcing prosecutors, regulators, and compliance professionals to rethink fundamental assumptions about health care fraud. When machines make decisions that would be criminal if made by humans, who bears responsibility? How do you prove intent when the fraud emerges from patterns in training data rather than explicit programming? And what happens when AI independently discovers new ways to defraud government health care programs that its creators never imagined?

The AI Fraud Landscape

Health care fraud schemes potentially empowered by AI fall into two broad categories, each presenting unique challenges for enforcement and compliance.

Billing fraud includes upcoding, billing for services not rendered, and charging for medically unnecessary procedures. These schemes directly submit false claims to government programs. Under the False Claims Act (FCA),[2] the government need only prove "knowledge" that claims are false—including reckless disregard or deliberate ignorance. This lower intent threshold makes AI particularly dangerous in billing contexts.

Kickback schemes involve paying for patient referrals, in violation of the federal Anti-Kickback Statute (AKS),[3] the Eliminating Kickbacks in Recovery Act (EKRA),[4] or similar state laws. Violations of these statutes require proving a "knowing and willful" quid pro quo arrangement—a higher bar that traditionally requires showing both parties understood the corrupt bargain. AI can obscure these arrangements by making referral patterns appear to be objective clinical decisions.

What makes AI-powered fraud fundamentally different is how intent manifests in machine learning systems. Unlike traditional "symbolic" programming where fraud might be coded as explicit rules ("IF Medicare patient THEN add malnutrition diagnosis"), machine learning involves training models on data patterns.[5] Even intentionally biased systems maintain plausible deniability because you can't point to a specific line of code that commits fraud.

Consider two scenarios that illustrate this challenge. In the first, a health system trains its clinical documentation AI on cases that maximized reimbursement, selects features that correlate with higher payments, and rewards the model for "complete" documentation that happens to trigger higher billing rates. While avoiding explicit programming of fraud, every design choice points toward the same outcome. This represents intentional design with built-in deniability.

In the second scenario, a health system trains its AI solely on clinical best practices and patient outcomes. Through complex interactions in the data, the AI independently discovers that certain diagnosis combinations increase reimbursement and begins suggesting them more frequently. The fraud occurs in the absence of human intent—a

truly emergent pattern.

This distinction could matter enormously for prosecution and compliance. But here's the challenge: from the outside, both scenarios look identical. The AI produces the same fraudulent billing patterns. The code contains no smoking guns. Machine learning's inherent complexity provides cover for those who would deliberately design fraudulent systems.

Billing Fraud in the Age of AI

The Automated Upcoder

AI's pattern-recognition capabilities make it exceptionally effective at identifying opportunities to maximize reimbursement. Consider an AI system trained on millions of medical records that learns which secondary diagnoses increase Medicare payments. The system begins prompting physicians to document conditions like "mild protein-calorie malnutrition" in elderly patients—technically defensible diagnoses that significantly increase reimbursement but may not be clinically relevant to the patient's care.

The sophistication goes deeper. Modern AI can analyze subtle patterns humans might miss: which diagnosis combinations trigger higher payments, which payer-specific rules can be exploited, which documentation phrases avoid audits. An AI might discover that adding "chronic systolic heart failure" to certain patient profiles increases payment by $3,000 while rarely triggering review—then systematically suggest this diagnosis for borderline cases.

Medical Necessity Manipulation

More troubling are AI systems that identify patients for profitable—but unnecessary—procedures. A cardiac AI might flag patients as "high risk" based on algorithms that weight factors correlating with good insurance coverage more heavily than actual medical risk factors. The AI provides clinical-sounding justifications for cardiac catheterizations that could be managed with medication, generating substantial facility fees and professional charges.

The AI might learn that certain clinical presentations combined with specific insurance types rarely face prior authorization challenges. It then recommends aggressive intervention for these patients while suggesting conservative treatment for clinically similar Medicaid patients. The bias emerges from training data but results in systematic fraud and disparate patient treatment.

Ghost Billing and Service Inflation

AI systems with access to scheduling and billing data can identify patterns in services rarely audited. The system might auto-populate medical records with standard add-on procedures that typically accompany primary services, regardless of whether they were performed. Or it might suggest "incident to" billing opportunities that technically comply with regulations but push the boundaries of legitimate practice.

The Unique Danger of Billing Fraud AI

What makes AI-powered billing fraud particularly dangerous is the FCA's lower scienter requirement. The government doesn't need to prove that a provider intended to submit false claims—only that it knew the claims were false. This could encompass reckless disregard,[6] such as deploying an AI without attempting to understand how it makes coding decisions; deliberate ignorance[7] when refusing to investigate after revenue suspiciously increases; or—the government might argue—simply situations where a company consciously avoided implementing reasonable safeguards despite clear risks.

This means that even truly "autonomous" fraudulent conduct—where the AI teaches itself to act deceitfully—can potentially create liability. Arguing that you did not program a model to commit fraud may not work if it can be shown that you had a good idea of what your AI was doing. Health care organizations deploying AI for billing or coding must therefore maintain vigilant oversight, as technical ignorance does not necessarily shield against FCA liability.

The AI-Powered Kickback

The Robo-Referral Scheme

Traditional kickback schemes can be relatively straightforward: for example, a laboratory pays a physician for each patient referred. AI-powered kickbacks can be more subtle.

Imagine that a laboratory provides physicians with a "free" AI-powered diagnostic support platform. The lab also pays physicians a certain amount per patient for "AI training data"—supposedly detailed outcome reports that help improve the algorithm's accuracy. In reality, physicians provide minimal feedback through auto-populated forms that rarely meaningfully improve the AI.

Meanwhile, the AI systematically recommends that laboratory for testing, citing sophisticated clinical rationales. When questioned, physicians claim they're following objective AI recommendations. The lab maintains that it's paying for legitimate data services. The AI provides cover for what is functionally a traditional kickback scheme.

Referral Pattern Manipulation

The scheme's sophistication lies in how AI obscures the quid pro quo. The AI might recommend the partner lab for 85% of tests while suggesting competitors only for unusual tests they don't perform. It provides detailed clinical justifications that make each referral seem medically necessary. To maintain an appearance of objectivity, it occasionally recommends competitors for low-value tests. Most insidiously, it might adjust recommendations based on the physician's historical compliance with suggestions, learning which doctors follow its guidance.

The AI learns these patterns not from explicit programming but from training data that includes historical referral patterns, "preferred provider" flags, or operational efficiency metrics that favor the financial partner. The sophistication makes proving traditional kickback intent extremely challenging.

The Temporal Evolution Problem

A unique challenge with AI kickbacks is how liability evolves over time. Consider this progression: In month one, a lab provides an AI tool and begins paying for training data with no referral bias yet existing. By month six, the AI begins showing preference for the lab based on "integration efficiency." At month twelve, a clear pattern emerges—physicians using the AI refer 80% of tests to the paying lab. By month eighteen, both parties are aware of the pattern but continue the arrangement.

When does this become illegal? The initial arrangement may be innocent, but if the lab continues after discovering the pattern, it could be accused of acting with the required corrupt intent. Each payment and referral after awareness could be viewed as a new violation. The parties could then no longer claim ignorance of the quid pro quo effect, or so the argument would go. This temporal evolution creates complex questions about when knowledge transforms into intent, and when continuation of an originally innocent arrangement becomes criminal.

When AI Directs Payments

Even more problematic are systems where AI determines payments based on referral value. An AI system might increase "training data" payments to physicians who generate more referrals, identify which physicians respond to payment incentives and adjust accordingly, or create complex payment formulas that correlate with referral value while maintaining plausible legitimate factors.

This scenario may be harder to defend because the AI directly implements the problematic payment structure rather than just influencing clinical decisions. The algorithm literally operationalizes the quid pro quo, making it more difficult to argue the correlation is coincidental.

The Mens Rea Challenge: When Algorithms Can't Form Intent

The most fundamental challenge in prosecuting AI-facilitated fraud lies in establishing criminal intent. Fraud and kickback liability generally requires *mens rea*—a "guilty mind." But algorithms don't have minds.

The Intent Laundering Problem

Machine learning[8] (ML) inherently provides plausible deniability for intent. Unlike traditional programming where one might find code explicitly directing fraud, ML involves training data selection that influences outcomes, rewards functions that optimize for certain results, architectures choices that enable certain patterns, and features engineering that emphasizes certain factors.

Each choice can be defended as technically legitimate while collectively producing fraudulent outcomes. Did the developer who included "payer type" as a feature intend for the AI to discriminate against Medicaid patients, or were they simply providing comprehensive data? Did the executive who approved training on high-revenue cases intend to create an upcoding machine, or were they showcasing "successful" treatments?

This ambiguity is not a bug, but a feature, for those seeking to commit fraud. The layers of mathematical complexity between human decisions and fraudulent outcomes create multiple opportunities to claim ignorance or legitimate purpose.

How Courts Might Adapt

Several evolving legal doctrines could address the AI intent gap.

The doctrine of "willful blindness"[9] prevents health care organizations from escaping liability by deliberately avoiding knowledge of how their AI systems work. If a health system implements an AI tool that triples Medicare revenue without investigating why, prosecutors may argue that it consciously avoided discovering the fraud. The complexity of AI, the government may insist, doesn't excuse the duty to understand your tools.

Courts may also apply the "collective knowledge" doctrine, recognizing that organizations don't escape liability simply because no single person understood the full scheme.[10] When the data science team knows the AI weights financial factors, the finance team celebrates increased revenue, and the clinical team notices biased patterns, their collective knowledge could arguably establish organizational intent.

The concept of "reckless disregard"[11] could also factor into this context. The government might argue that deploying AI without understanding its decision-making process may constitute reckless disregard for truth or falsity, especially in billing fraud cases under the FCA where such reckless disregard suffices for liability.

The Human Element

Importantly, for a fraud or kickback case, it does not suffice (and is impossible) to prove that the AI itself had intent. Rather, prosecutors or whistleblowers would need to prove that relevant *humans* had the requisite intent or knowledge in designing, deploying, or continuing to use the AI despite fraudulent outcomes.

Evidence might include emails discussing how to "optimize" the AI for revenue, training data selections that predictably lead to fraud, failure to implement standard bias testing, continuation of AI use after problematic patterns emerge, or rejection of safeguards that would prevent fraudulent outcomes. The focus remains on human choices about the AI, not the AI's "choices" themselves.

Enforcement Evolution

The Three-Pronged Investigation Approach

Law enforcement is adapting to AI fraud with sophisticated investigative strategies that combine traditional and novel approaches.

Traditional evidence gathering remains crucial. Prosecutors will seek communications about AI design decisions, including emails discussing training data selection, meeting notes about model optimization goals, Slack messages joking about the AI's revenue generation, and board presentations touting AI-driven income increases. These human communications can reveal intent even when the code doesn't.

Technical forensics represents a new frontier in health care fraud investigation. Expert witnesses will reverse-engineer AI systems to demonstrate bias through analysis of training data composition, examination of feature weights and model architecture, testing with synthetic patients to reveal discrimination, and comparison with legitimate clinical AI systems. This technical analysis, on which researchers have been making progress,[12] can show how design choices predictably led to fraudulent outcomes.

Statistical pattern analysis provides powerful circumstantial evidence. Prosecutors will seek to present referral patterns that couldn't occur naturally, billing distributions that differ dramatically from peers, correlation between AI recommendations and financial relationships, and timeline analysis showing changes after AI deployment. These patterns could be so striking that they speak for themselves, even without traditional evidence of agreements.

Evidentiary Challenges in the AI Era

These new investigative methods raise issues as to how courts will handle AI-generated evidence and analysis. As prosecutors and whistleblowers increasingly rely on sophisticated technical analysis to prove fraud schemes, courts must grapple with admissibility standards, authentication requirements, and the need for specialized expertise.

In recognition of these challenges, the Advisory Committee on Evidence Rules has proposed a new Federal Rule of Evidence 707.[13] The proposed rule would subject machine-generated evidence offered without an expert witness to the same *Daubert* standards traditionally applied to expert testimony, requiring proof that the evidence is based on sufficient facts, uses reliable principles and methods, and has been reliably applied to the case at hand.

Beyond Rule 707, proposed amendments to Rule 901[14] would create specific requirements for authenticating AI-generated evidence. The party seeking to introduce the evidence would need to introduce foundational evidence that describes the training data and model used and shows that it produced reliable results.

The New Whistleblowers

Data scientists and IT professionals may become the new health care fraud whistleblowers. They understand how algorithms can be manipulated and can provide technical evidence of bias. A data scientist who discovers their employer trained an AI on revenue-maximizing cases, or who is asked to adjust algorithms to increase referrals, could have a potential qui tam case.

The FCA's whistleblower provisions are particularly powerful here because they protect and incentivize insiders who understand the technical details. We're likely to see cases brought by data scientists asked to implement problematic features, compliance officers who discover AI bias, physicians who notice systematic AI-driven fraud, and competing companies that identify suspicious patterns through market analysis. Congress is currently considering a new statute—the AI Whistleblower Protection Act[15]—aimed at preventing retaliation against such potential qui tam relators.

Evolving Prosecutorial Tactics

The Department of Justice (DOJ) has signaled its focus on AI-facilitated health care fraud through initiatives examining AI's role in fraud schemes. Prosecutors are developing new strategies that reflect the unique challenges of AI-driven fraud. These include using AI to analyze claims data and identify AI-driven fraud patterns—fighting fire with fire.

Over the past year, DOJ has begun issuing subpoenas to pharmaceutical and digital health companies regarding AI tools embedded in electronic health records that prompt doctors to recommend specific treatments.[16] It has pursued cases against Medicare Advantage plans for using algorithms designed to identify diagnosis codes that increase revenue while failing to implement corresponding algorithms to identify and correct inaccurately reported codes. And its September 2024 memorandum on Evolution of Corporate Compliance Programs[17] directed prosecutors, in evaluating a company's compliance efforts, to consider its policies regarding identifying and mitigating risks of misconduct resulting from use of AI.

Regulatory and Legislative Response

Emerging AI Governance Requirements

States like Colorado[18] and jurisdictions like the EU[19] are pioneering AI governance laws that could reshape health care AI compliance. For health care AI, we can expect comprehensive new requirements for consequential decisions involving high-risk processing.

Mandatory bias testing would require regular audits comparing AI recommendations against clinical benchmarks, with documentation that referral patterns stem from medical necessity rather than financial relationships. Organizations would need to show not just that their AI works, but that it works fairly and without financial bias.

Transparency requirements would mandate explainable AI for health care decisions, allowing users and auditors to understand why specific recommendations were made. Audit trails would have to show recommendation rationales in clinical terms. Organizations would need to disclose training data sources and any financial relationships that could influence AI behavior. Regular reporting of outcome patterns would allow regulators to identify systematic bias across the industry.

Before deployment, health care AI could have to undergo rigorous impact assessments testing for discriminatory patterns, financial bias, and potential for fraudulent outcomes. These assessments would be updated regularly as the AI learns and evolves and would provide a useful mechanism for tracking model drift.

Continuous monitoring obligations could require organizations to detect and remediate emerging bias in real-time, with requirements to report problematic patterns to regulators. This would shift compliance from a point-in-time assessment to an ongoing obligation. To that end, many of the emerging AI governance laws require companies to implement an AI governance program with ongoing monitoring and safety testing to proactively prevent algorithmic discrimination.

Health Care–Specific Regulations

Beyond general AI governance, we might see health care-specific rules that address the unique risks in medical AI.

Clinical validation requirements could mandate that AI demonstrate clinical efficacy independent of financial outcomes. Organizations would have to show medical benefit through rigorous studies before considering any revenue impact. This would reverse the current trend, in which financial success often drives adoption.

Financial firewall mandates would require separation between clinical AI development and revenue cycle teams. Organizations might have to document that clinical decisions are not influenced by financial data, creating clear boundaries between care and commerce. Because patient care should never be dictated by corporate profits, it is crucial to align models with the goal of improved patient care rather than increased revenue.

For high-risk decisions such as expensive procedures or specialty referrals, regulations may require human-in-the-loop review with documentation of clinical rationale. This would ensure a licensed professional takes responsibility for significant medical decisions.

Regulators might also create safe harbors protecting organizations that follow prescribed compliance protocols. This carrot-and-stick approach would incentivize proactive bias prevention while punishing willful blindness. For example, the Colorado AI Act provides an affirmative defense provision for companies that have implemented National Institute of Standards and Technology (NIST) or International Organization for Standardization (ISO) frameworks, self-report harms, and are otherwise in compliance with the statute.

Compliance Best Practices

Health care organizations can take concrete steps to prevent AI from becoming a vehicle for fraud, or at least from inspiring suspicion from the wrong quarters.

Document Design Principles and Decisions

Organizations must train models exclusively on clinical outcomes and evidence-based guidelines while rigorously excluding from training sets any improper revenue, reimbursement, or referral data that could introduce improper bias. Every data point should have a documented clinical justification. When the finance team wants to add payer information to the model, relevant decision makers should be empowered to resist where necessary.

Structural safeguards can prevent concentration of referrals even when the AI makes clinically sound recommendations. When multiple providers are clinically equivalent, the system could randomize selection rather than developing patterns that could appear suspicious. Organizations could establish concentration thresholds that trigger alerts when any facility receives disproportionate referrals. Human approval could be required for referrals to financially connected entities, creating a checkpoint where potential conflicts are acknowledged and justified.

The principle of transparent architecture calls for AI systems that can explain their recommendations in clinical terms. Under this principle, every decision should be auditable back to specific training data and model features. This isn't just about satisfying regulators—it's about maintaining physician trust and ensuring medical integrity.

Implement Frameworks and Controls

Pre-deployment testing must verify that AI makes decisions based on clinical factors. Organizations can test AI with synthetic patients across different payer types and demographic groups. Results could then be compared to historical baselines and peer benchmarks. Any financial benefits from AI implementation could be scrutinized for improper patterns.

Ongoing monitoring could involve sophisticated analytics examining referral and billing patterns monthly, comparing results with pre-AI baselines and regional norms, tracking denial rates and audit results, and regularly reviewing whether "AI training data" payments actually improve model performance. This monitoring would ideally be independent of the teams that benefit from AI-driven revenue.

The best tool for managing oversight and documenting model performance is to implement an internationally recognized guideline like the NIST AI Risk Management Framework or the ISO 42001 Framework. These tools provide a structure for tracking, monitoring, testing, and improving model performance.

Human oversight remains essential despite AI sophistication. Organizations should establish clear thresholds for human review based on the dollar value of services recommended, deviation from typical treatment patterns, recommendations favoring financially connected entities, and unusual diagnosis combinations. The goal is not to second-guess every AI decision but to catch patterns that might indicate bias. Ongoing monitoring and safety testing with a human-in-the-loop would be a best practice.

Documentation Requirements

Comprehensive documentation serves both compliance and defense purposes. Development records should capture training data selection rationale, model architecture decisions, testing protocols and results, and clinical validation studies. This contemporaneous documentation can prove good faith if problems later emerge. Documenting model testing and oversight will provide evidence of intent to comply with statutory requirements in the event that a company is confronted with a regulatory investigation.

Operational logs should track recommendation patterns by provider and payer, override rates and justifications, updates and their clinical rationales. Financial impact analyses should be kept separate from clinical operations. Historical logs with safety testing, well-reasoned decisions, harm identification, and documented remedial actions can help demonstrate that any suspicious patterns were identified and addressed.

Compliance documentation should include regular audit results, remediation efforts when patterns are detected, board and committee oversight minutes, and training provided to users. This creates a paper trail showing the organization took its obligations seriously.

When Problems Emerge

Despite best efforts, problematic patterns may emerge. The organization's response can mean the difference between a correctable error and criminal liability.

Upon discovering bias, organizations must act immediately. Document the finding and begin remediation without delay. Continuing to use biased AI after discovery dramatically increases liability and undermines any defense based on lack of intent.

Investigation must be thorough and honest. Determine whether bias was intentionally designed or truly emergent. Review all development documentation and interview key personnel. Engage outside experts if necessary to ensure objectivity.

Remediation options include retraining models without biased data, implementing additional human oversight, adding randomization or bias-correction features, or temporarily suspending AI use for affected decisions. The key is showing that the organization takes the problem seriously and is committed to fixing it. Thorough documentation, implementation of a recognized framework, and evidence of good faith are the best defense to a regulatory investigation related to emerging technologies.

Conclusion

AI-powered health care fraud represents an evolution, not a revolution, in illegal schemes. The underlying crimes—paying for referrals and billing for unnecessary services—remain unchanged. What's new is the sophistication of the concealment mechanism and the scale at which fraud can occur.

Health care organizations must recognize that deploying AI doesn't eliminate compliance obligations—it may actually amplify them. The same technology that can improve patient care can also systematically defraud health care programs at unprecedented scale. The difference lies not in the technology itself but in how it's designed, implemented, and monitored.

The era of "the algorithm did it" as a defense is ending before it truly began. Courts and prosecutors will be adapting existing legal doctrines to address AI accountability gaps. Regulators are implementing new frameworks requiring transparency, testing, and continuous monitoring. Whistleblowers with technical expertise are identifying AI-driven fraud patterns.

For health care organizations, the message is clear: invest in compliance infrastructure now or face potentially serious liability later. This means building AI systems with clinical integrity, implementing robust monitoring, maintaining comprehensive documentation, and acting swiftly when problems emerge.

The future of health care AI can be bright—improving diagnosis accuracy, treatment selection, and operational efficiency—but only if we prevent it from becoming a tool for laundering fraudulent intent through mathematical complexity. In this new landscape, responsible AI deployment isn't just about avoiding liability; it's about maintaining the integrity of our health care system and the trust of the patients it serves.

About the Authors

Joshua Robbins is a former federal health care fraud prosecutor and co-chair of the White Collar & Investigations practice at Buchalter. **Daniel Pietragallo** is a former Senior Assistant Attorney General for Colorado and co-chair of the firm's Artificial Intelligence practice.

---

[1] Thomas Woodside, *Emergent Abilities in Large Language Models: An Explainer*, Ctr. for Security and Emerging Tech., Apr. 16, 2024, https://cset.georgetown.edu/article/emergent-abilities-in-large-language-models-an-explainer/.

[2] 31 U.S.C. § 3729 et seq.

[3] 42 U.S.C § 1320a-7b.

[4] 18 U.S.C § 220.

[5] Neeraj Kumar and Rahul Rai, *What is the difference between symbolic systems and machine learning?*, AIBRILLANCE, May 22, 2004, https://www.aibrilliance.com/blog/what-is-the-difference-between-symbolic-systems-and-machine-learning.

[6] *See* Manual of Model Civ. Jury Instructions, Knowledge, https://www.ce9.uscourts.gov/jury-instructions/node/293.

[7] *See* Manual of Model Crim. Jury Instructions, Deliberate Ignorance, https://www.ce9.uscourts.gov/jury-instructions/node/859.

[8] Sara Brown, *Machine learning, explained*, MIT Sloan Sch. of Mgmt., Apr. 21, 2021, https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained.

[9] *See* Manual of Model Crim. Jury Instructions, Deliberate Ignorance, https://www.ce9.uscourts.gov/jury-instructions/node/859.

[10] *United States v. Bank of New England*, 821 F.2d 844 (1st Cir. 1987), https://law.justia.com/cases/federal/appellate-courts/F2/821/844/255827/.

[11] *See* Manual of Model Civ. Jury Instructions, Knowledge, https://www.ce9.uscourts.gov/jury-instructions/node/293.

[12] Billy Perrigo, *No One Truly Knows How AI Systems Work. A New Discovery Could Change That*, TIME, May 21, 2024, https://time.com/6980210/anthropic-interpretability-ai-safety-research/.

[13] Proposed Amendments, Fed. R. Evid. 707 (June 10, 2025), https://www.uscourts.gov/sites/default/files/document/2025-06-standing-agenda-book.pdf.

[14] Proposed Amendments, Fed. R. Evid. 901 (June 10, 2025), https://www.uscourts.gov/sites/default/files/document/2025-06-standing-agenda-book.pdf.

[15] AI Whistleblower Protection Act, S.1792, 119th Cong. (2025-2026), https://www.congress.gov/bill/119th-congress/senate-bill/1792/text.

[16] Ben Penn, *DOJ's Healthcare Probes of AI Tools Rooted in Purdue Pharma Case*, BLOOMBERG, Jan. 29, 2024, https://news.bloomberglaw.com/product/blaw/bloomberglawnews/exp/eyJpZCI6IjAwMDAwMThkLTM2ZTEtZGVi source=newsletter&item=body-link&region=text-section.

[17] U.S. Dep't of Justice, Crim. Div., Evaluation of Corporate Compliance Programs (Updated Sept. 2024), https://www.justice.gov/criminal/criminal-fraud/page/file/937501/dl?inline=.

[18] SB 24-205, https://leg.colorado.gov/bills/sb24-205.

[19] EU Artificial Intelligence Act, https://artificialintelligenceact.eu/ai-act-explorer/.

ARTICLE TAGS

Fraud and Compliance | Health Information