S O P H I O N . A I

Trust Infrastructure for AI in Healthcare:

A Framework for Accountable, Auditable, and Ethical Autonomous Clinical AI Systems

A Response to the HHS Request for Information:
Accelerating the Adoption and Use of Artificial Intelligence

as Part of Clinical Care (90 FR 60108)

David Gage

Founder & CEO, SophionAI
February 2026

Executive Summary

The U.S. Department of Health and Human Services has identified a critical truth: the next phase of AI in healthcare will be driven not by model capability alone, but by trust, accountability, and policy alignment. SophionAI agrees. We are building the trust infrastructure that makes autonomous AI in clinical settings governable, auditable, and safe.

This paper responds to the HHS Request for Information (90 FR 60108) from the perspective of a company that operates at the intersection of clinical operations, AI systems engineering, and healthcare ethics. Our founder brings over twenty years of direct healthcare administration experience, from managing a family practice to establishing a psychiatry practice, providing firsthand knowledge of the regulatory, billing, and documentation burdens that AI must address responsibly.

Our central thesis: the healthcare industry does not lack AI models. It lacks the governance layer that makes those models trustworthy enough to act autonomously in regulated environments. Until that layer exists, AI adoption in clinical care will remain slow, fragmented, and vulnerable to the same trust failures that have plagued health IT adoption for decades.

> *We do not argue for faster AI deployment. We argue for trust infrastructure that makes faster deployment safe.*

1. The Trust Gap in Healthcare AI

Healthcare AI faces a fundamental credibility problem. Clinicians, administrators, and regulators are being asked to trust systems that cannot explain their reasoning, cannot prove their decisions were made ethically, and cannot be audited after the fact. This is not a technology problem. It is a governance problem.

Today, most AI tools deployed in clinical settings operate as black boxes. A model produces a recommendation. The clinician either accepts or rejects it. If something goes wrong, there is no audit trail, no provenance chain, no way to reconstruct why the system said what it said. This is unacceptable in an industry where documentation is the legal record, where billing codes carry fraud liability, and where patient safety is paramount.

The HHS RFI asks what barriers slow AI adoption. From our operational experience, the answer is not cost or technology. It is trust. Providers do not trust that AI tools will protect them. Administrators do not trust that AI outputs are auditable. And regulators do not yet have frameworks to evaluate whether an AI system meets the standard of care expected of the humans it assists.

What Trust Infrastructure Requires

Trustworthy AI in healthcare demands several properties that current systems largely lack. Every AI-generated recommendation, score, or action must be accompanied by an auditable

record of how it was produced, what data informed it, what ethical considerations were evaluated, and what confidence level the system assigned. This audit trail must be tamper-evident, meaning any alteration to the historical record is detectable. It must be reproducible, meaning an auditor can reconstruct the decision with the same inputs, models, and thresholds that existed at the time. And it must separate what the AI learned from whose data it learned from, preserving patient privacy at the architectural level rather than as a compliance afterthought.

2. The Broken Data Foundation

The HHS RFI correctly identifies interoperability as a key lever for AI adoption. But the problem goes deeper than standards and data exchange protocols. The fundamental data infrastructure that AI systems must work with is broken.

Electronic Health Record systems were designed for billing and documentation, not for the kind of structured data access that AI requires. EHR data exports produce fragmented files that are frequently missing critical identifiers. Patient records that appear complete in the user interface become fractured, incomplete datasets when extracted through standard APIs. Rebuilding complete clinical pictures from these exports requires expensive manual reconstruction that negates the efficiency gains AI is supposed to deliver.

This is not an abstract concern. It is a daily operational reality in clinics and hospitals across the country. Any AI strategy that assumes clean, structured, interoperable data as a starting point is building on a foundation that does not exist for the vast majority of healthcare providers, particularly independent practices and small group settings.

A Practical Alternative

At SophionAI, we have adopted an approach that works with the healthcare data environment as it actually exists, rather than as we wish it were. Rather than relying on broken export pipelines, our system interacts with EHR platforms the same way a trained human employee would: through the user interface, with proper credentials, under the same access controls and audit logging that apply to any staff member. This approach provides access to the complete clinical picture that exists in the system, not the fragmented version that exports produce.

This is not a workaround. It is a design philosophy. AI systems that can only operate on exported, structured data will always be limited by the quality of that export pipeline. Systems designed to work within the existing clinical workflow, using the same interfaces clinicians use, can deliver value immediately without waiting for the interoperability problem to be solved industry-wide.

3. Ethical Governance as Architecture, Not Policy

The healthcare AI industry frequently discusses ethics in terms of policy documents, review boards, and compliance checklists. These are necessary but insufficient. Ethics must be embedded in the system architecture itself, not layered on top after the system is built.

SophionAI operates under the Sophionic Oath, a seven-point ethical framework that governs every architectural decision, every line of code, and every design tradeoff in our system. The oath establishes principles that are not aspirational statements but engineering constraints.

Principle	Architectural Implication
First, Do No Harm	AI must never be weaponized against providers, patients, or healthcare workers. The system serves as a guide, not a judge; a mentor, not an enforcer.
Support, Not Punishment	Systems must offer pathways for mentorship and remediation before any action that could negatively impact a provider. Scoring systems measure documentation quality to help providers improve, never to penalize.
Protecting the Vulnerable	When children or elderly patients are at risk, the system escalates to a human decision-maker. AI flags concerns but never acts as unchecked authority.
Privacy as Architecture	Patient privacy is enforced through governed access, not avoidance. The system handles protected health information with auditable controls, tamper-evident logging, and governance rigor that exceeds what human workflows typically provide. Every interaction with patient data is bounded, logged, and accountable.
Transparency	Every AI-generated score, flag, or recommendation includes a structured rationale. No black-box decision-making. Every decision is auditable and challengeable.
Responsible Deployment	Careful selection of who has access and how technology is deployed. The system cannot be exploited for corporate or regulatory interests at the expense of fairness.
Continuous Reflection	The ethical framework itself is versioned and evolves. As AI capabilities grow, governance thresholds and risk assessments are recalibrated against real-world outcomes.

When we say ethics is architecture, we mean that our system enforces governance constraints that human workflows routinely fail to uphold. In every healthcare organization, human employees access patient data daily. Some handle it responsibly. Others gossip, share information inappropriately, blackball patients, or fail to follow HIPAA protocols. The accountability mechanisms are inconsistent at best. Our system handles patient data with the same access a trusted employee would have, but with governance that no human workforce can match: every access is logged immutably, every decision is auditable, every interaction with protected health information is bounded by architectural constraints that cannot be overridden by convenience or negligence. The question is not whether AI should handle PHI. It is whether we can build AI systems that handle it more responsibly than the status quo. We believe we can, and the governance infrastructure to prove it is what makes the difference.

4. Responding to the HHS RFI: Specific Recommendations

The following recommendations respond to specific questions posed in the HHS RFI, informed by our experience building trust infrastructure for autonomous clinical AI and by over two decades of operational healthcare administration.

On Barriers to AI Adoption (Question 1)

The primary barrier is not technology or cost. It is the absence of a recognized standard for AI accountability in clinical settings. Providers and health systems lack a framework for evaluating whether an AI tool meets the governance requirements that their liability exposure demands. Without such a framework, every adoption decision is a bespoke risk assessment, which is prohibitively expensive for independent practices and small groups that stand to benefit most from AI-driven efficiency.

We recommend that HHS support the development of industry-driven accountability standards for clinical AI that address auditability (can every AI decision be reconstructed?), transparency (does the system explain its reasoning?), privacy architecture (is patient data protected structurally, not just procedurally?), and human oversight (are there defined escalation paths for high-risk decisions?).

On Regulatory Design (Question 2)

Current regulatory frameworks struggle with AI because they were designed for static software, not learning systems. An AI tool that improves over time through validated learning does not fit neatly into the traditional software certification model. We recommend that HHS explore a governance certification approach that evaluates not just the AI model itself, but the trust infrastructure surrounding it: how decisions are logged, how learning is validated, how patient data boundaries are enforced, and how human oversight is maintained.

This approach would allow HHS to regulate the governance layer rather than the model layer. Models change frequently. Governance frameworks, if well-designed, provide stable accountability regardless of which model is underneath.

On Novel Legal and Implementation Issues (Question 3)

The most pressing novel issue is liability allocation when AI systems operate with increasing autonomy. If an AI system reviews a patient chart, identifies a documentation gap, and sends a suggestion to a provider, who is liable if the suggestion is wrong? If the suggestion is right but the provider ignores it, does the existence of the AI suggestion change the provider's liability?

These questions cannot be answered in the abstract. They require that AI systems produce tamper-evident audit trails showing exactly what the system recommended, when, based on what evidence, and with what confidence level. HHS can help by establishing minimum audit trail requirements for AI systems operating in clinical settings, providing the evidentiary foundation that liability frameworks need.

On Evaluation Methods (Question 4)

Pre-deployment evaluation of clinical AI is necessary but insufficient. The greater challenge is post-deployment monitoring: ensuring that an AI system continues to perform as expected as clinical workflows change, EHR systems update, and the population it serves evolves. We recommend that HHS support the development of continuous evaluation frameworks that include ongoing documentation quality scoring, drift detection in AI outputs, and human validation sampling.

At SophionAI, our system runs self-maintenance cycles that detect when the clinical environment has changed (for example, when an EHR vendor updates its interface) and either adapts automatically for minor changes or halts operations and escalates for significant changes. This kind of continuous self-evaluation should be a standard expectation for clinical AI systems, not an optional feature.

On Supporting Private Sector Activities (Question 5)

HHS can accelerate AI adoption most effectively by supporting the emergence of industry-driven trust certification programs. Rather than certifying AI models directly (which is impractical given the pace of model development), HHS could recognize private sector certification bodies that evaluate the governance infrastructure surrounding clinical AI: audit trail integrity, privacy architecture, ethical gate mechanisms, and human oversight protocols.

This approach mirrors how the financial industry handles audit standards: the government sets expectations, and accredited private bodies perform the evaluations. Applied to healthcare AI, this would give providers and health systems a recognizable trust signal when evaluating AI tools, dramatically reducing the bespoke risk assessment burden that currently slows adoption.

On Interoperability (Question 8)

Enhanced interoperability is essential, but the industry should not wait for perfect interoperability to deploy AI. The reality is that interoperability improvements will take years. In the interim, AI systems must be designed to work with the clinical data environment as it exists today, including fragmented EHR exports, inconsistent data standards, and siloed information systems.

We recommend that HHS encourage the development of AI systems that are EHR-agnostic by design, capable of operating across different platforms without requiring custom integration for each system. This reduces deployment friction and ensures that AI benefits are not limited to health systems with the resources for custom technical integration.

On Research Priorities (Question 10)

We recommend that HHS prioritize research in three areas. First, trust and governance frameworks for autonomous AI agents in clinical settings, including audit trail standards, tamper-evidence mechanisms, and privacy-preserving learning architectures. Second, evaluation methodologies for AI systems that learn and improve over time, as opposed to static software. Third, human-AI collaboration models that maintain meaningful human oversight without negating the efficiency gains that AI provides.

5. The Case for Administrative AI

Much of the public conversation about AI in healthcare focuses on clinical decision support: helping doctors diagnose, helping radiologists read images. These are important applications, but they represent only a fraction of the opportunity.

The larger and more immediate opportunity is administrative AI: systems that reduce the documentation burden, optimize billing accuracy, ensure compliance, and streamline the operational workflows that consume an enormous portion of healthcare spending. Administrative burden is not a minor inconvenience. Studies consistently show that physicians spend more time on documentation and administrative tasks than on direct patient care. This burden drives burnout, reduces the time available for patients, and creates operational inefficiency that inflates costs across the system.

AI systems focused on documentation quality, billing optimization, and administrative workflow automation can deliver measurable value immediately, without the clinical liability concerns that slow adoption of diagnostic AI. A system that identifies a missing medication reconciliation in a chart, or flags an under-coded billing encounter, or detects a compliance gap in documentation, is operating in the administrative domain where the standard of care is well-defined and the risk profile is manageable.

SophionAI is purpose-built for this domain. Our system supports documentation quality and reduces administrative friction. It does not make clinical diagnoses or treatment recommendations. This scope discipline is intentional: it allows us to deliver immediate value while operating within a well-understood regulatory and liability framework.

6. Payment Models and AI Incentive Alignment

The HHS RFI correctly identifies that legacy payment systems can diminish the potential of AI innovations. We see this directly in the value-based care environment. Accountable Care Organizations and similar risk-bearing entities have a financial incentive to invest in documentation quality and preventive care optimization, because capturing accurate risk adjustment and incentive payments depends on complete, well-coded clinical documentation.

AI systems that improve documentation quality and billing accuracy can generate measurable financial returns in this environment, because the gap between what providers earn and what they could earn with complete, accurate documentation is substantial. In many independent practices, this gap represents tens of thousands of dollars annually in uncaptured quality incentive payments alone.

We recommend that HHS consider payment policy changes that explicitly recognize AI-driven documentation quality improvement as a reimbursable activity, particularly in value-based care programs where documentation accuracy directly affects payment. This would create a virtuous

cycle: better documentation drives better risk adjustment, which funds further investment in AI tools, which drives better documentation.

7. Preserving Independent Practice

AI adoption in healthcare is at risk of following the same consolidation pattern that has characterized health IT over the past two decades: large health systems with dedicated IT departments adopt early, while independent practices and small groups are left behind or forced into acquisitions by larger entities that can afford the technology.

This is not just a business concern. Independent practices provide a disproportionate share of primary care, particularly in rural and underserved communities. If AI-driven efficiency becomes available only to large consolidated systems, the competitive pressure on independent practices will accelerate consolidation, reducing patient choice and access in the communities that need it most.

SophionAI is designed specifically to serve independent practices and small groups. Our system operates on commodity hardware, runs locally with minimal cloud dependency, and does not require dedicated IT staff to maintain. We believe HHS should consider the impact on independent practice viability when evaluating AI policy, and should ensure that AI adoption incentives do not inadvertently favor large health systems at the expense of the independent providers who serve our most vulnerable communities.

8. Conclusion

The HHS Request for Information represents a pivotal moment. The Department is asking the right questions: not just how to make AI faster, but how to make it trustworthy, accountable, and aligned with the values that healthcare demands.

SophionAI believes the answer lies in trust infrastructure. Not trust as a marketing term, but trust as an engineering discipline: auditable decision trails, tamper-evident records, privacy-preserving learning, ethical governance embedded at the architectural level, and meaningful human oversight that scales. We are building this infrastructure, and we welcome the opportunity to contribute to the national conversation about how AI can serve healthcare providers, patients, and communities responsibly.

> *First, do no harm. In the age of autonomous AI, that principle demands not just intention, but architecture.*

About SophionAI

SophionAI is a Texas-based AI company building trust infrastructure for autonomous AI in healthcare. Founded by David Gage, who brings over twenty years of direct healthcare administration experience, SophionAI develops systems that support clinical documentation quality, billing optimization, and administrative workflow automation for independent medical practices and healthcare organizations. The company operates under the Sophionic Oath, a seven-point ethical framework that governs all technical and business decisions.

Contact: David Gage, Founder & CEO | SophionAI | sophion.ai