A PRAXISONE PERSPECTIVE

Governed by Design

Responsible AI in behavioral health — and why governance is the precondition for adoption, not a brake on it.

Executive summary

Artificial intelligence is entering behavioral healthcare at precisely the moment the field can least afford to get it wrong: the most sensitive data in medicine, some of its most vulnerable patients, and safety decisions where an error is not a bug ticket but a harm. At the same time, the oversight around health AI is tightening quickly — the FDA's line between decision support and regulated device, new federal transparency requirements for predictive tools, state laws such as the Texas Responsible AI Governance Act, and voluntary but fast-adopting standards from the Joint Commission and the Coalition for Health AI. The organizations that succeed with AI in behavioral health will not be the ones that move fastest or slowest, but the ones that move *under governance*. This paper maps the tightening landscape and lays out a practical model for governing behavioral health AI — human oversight, transparency, monitoring, fairness, data protection, and accountability — and argues that responsible governance is not a brake on adoption. In a field this sensitive, it is the precondition for adoption at all.

1 The highest-stakes place to deploy AI

Every use of AI in medicine carries responsibility. Behavioral health concentrates it. The data is among the most sensitive that exists — protected not only by HIPAA but by the additional wall of 42 CFR Part 2 for substance-use records. The patients are often vulnerable and in crisis. And the decisions AI touches — risk, safety, disposition — are precisely the ones where a wrong or blindly-trusted output can lead directly to harm.

This is also the category regulators watch most closely. Predictive, adaptive models that score patient risk sit at the center of the FDA's attention, because they are the tools most likely to cross the line from *supporting* a clinician's judgment to *substituting* for it. In behavioral health, getting AI governance wrong is not a compliance footnote. It is a patient-safety and liability event waiting to happen.

2 The oversight web is tightening

An organization adopting AI in behavioral health today is stepping into a landscape that did not exist three years ago — a fast-converging web of federal rules, state laws, and voluntary standards. Four reference points define its current shape.

The FDA distinguishes decision-support software that is a regulated medical device from software that is not. The pivotal criterion is independence: to stay on the non-device side of the line, the software must

let a clinician *independently review the basis* for its recommendation — the framework's explicit guard against a provider relying primarily on the machine. Behavioral health's adaptive risk models live closest to that line.

Federal transparency rules now reach predictive tools directly.

> **31 source attributes**
>
> Federal health-IT rules (the ONC HTI-1 Final Rule) require predictive decision-support tools to disclose 31 "source attributes" — in effect a nutrition label for the algorithm — to the people who use them, with intervention risk management around validity, fairness, and safety. In force since 2025.
>
> *ONC / HHS, Health Data, Technology, and Interoperability (HTI-1) Final Rule, 2024*

State law is moving faster still, and Texas is at the front.

> **$10,000–$200,000**
>
> per violation are the civil penalties under the Texas Responsible AI Governance Act (TRAIGA), effective January 1, 2026. Among its requirements: health care providers must clearly and conspicuously disclose their use of AI in treatment. Enforcement rests with the Texas Attorney General.
>
> *Texas Responsible Artificial Intelligence Governance Act (TRAIGA), 2026*

And **voluntary standards** are quickly becoming the expected baseline.

> **7 elements**
>
> The Joint Commission and the Coalition for Health AI (CHAI) issued 2025 guidance defining seven elements of responsible AI use: governance structures, patient transparency, data security, ongoing quality monitoring, blinded safety-event reporting, bias assessment, and staff education.
>
> *The Joint Commission & CHAI, Responsible Use of AI in Healthcare, 2025*

Underneath all of it sits the **NIST AI Risk Management Framework** — its four functions, *Govern*, *Map*, *Measure*, *and Manage*, now the common vocabulary for AI risk across industries. The through-line of every one of these is the same: transparency, human oversight, monitoring, fairness, and accountability. Whether or not a given tool is formally "regulated," those expectations are converging into table stakes.

3 Governance is not a brake — it is the enabling condition

It is tempting to read a landscape like that as friction: a thicket of rules that slows innovation down. In behavioral health, the opposite is true. Ungoverned AI in this field is not fast — it is unadoptable.

A clinician will not trust a black-box score they cannot question. A patient will not consent to a tool no one will explain. A regulator, an auditor, or a plaintiff's attorney will not accept "the algorithm decided" as a defense. Every stakeholder whose buy-in an organization needs — clinical, legal, regulatory, and the patient themselves — requires the very things governance provides: explainability, oversight, a record. Governance is not the tax you pay to use AI in behavioral health. It is the thing that makes using it possible.

4 A practical governance model for behavioral health AI

The frameworks above converge on a workable model. Six commitments turn “responsible AI” from a slogan into an operating discipline:

- **Human oversight, independently reviewable.** Every AI output can be traced to its basis and weighed by a licensed professional, who makes the decision. The system informs; it never acts on a patient on its own. This is the FDA's independence principle, built in rather than bolted on.
- **Transparency.** The organization knows — and can disclose — what each tool is, how it performs, and where its limits are, and it discloses the use of AI in treatment where law requires. This is the federal “source attributes” and the state disclosure duty, met by design.
- **Continuous monitoring and validation.** Tools are tested before deployment and re-validated after, with active watch for performance drift. A model that was accurate at launch is not assumed to stay that way.
- **Bias and fairness assessment.** Tools are evaluated, before and after go-live, for disparate impact across the populations they touch — with special care for vulnerable groups.
- **Data protection.** HIPAA and 42 CFR Part 2 are honored; protected health information is not used to train shared models; re-identification is prohibited; access is role-based and logged.
- **Governance structure and accountability.** A named, multidisciplinary oversight body owns AI policy; safety events are reported and learned from; staff are trained. Someone is accountable, by name.

5 Build it in, don't bolt it on

The single most important word in that model is *design*. Governance that is added at the end — a policy binder written after the tool is live, a disclosure bolted onto a product that cannot actually explain itself — satisfies no one and protects nothing. The auditable trail, the explainable output, the human decision point, the data segregation: these either exist in the architecture or they do not.

That is the real dividing line between organizations, and between the vendors they choose. Some treat governance as a filing to be produced under pressure. Others treat it as a feature — engineered in, demonstrable on demand, and ready for oversight that has not fully arrived yet. In a field where the rules are still tightening, only the second kind is safe to build on.

6 A readiness checklist for boards and compliance leaders

Leaders can pressure-test their organization's — or a vendor's — readiness with a short set of questions:

- Do we have a **named, multidisciplinary AI governance body**, with real authority and clear ownership?
- Can **every AI output be independently reviewed** to its basis by the clinician using it?
- Do we **disclose our use of AI** to patients wherever law and good practice require?
- Do we **monitor for drift and bias** after deployment — not just validate once at launch?
- Is **PHI protected** under HIPAA and 42 CFR Part 2, kept out of shared model training, and shielded from re-identification?

- Can we **produce an audit trail** — every signal, action, owner, and time — on demand?

7 Trust is the platform

In behavioral health, trust is not a feature of the product. It *is* the product. A patient hands over the most private facts of their life; a clinician stakes their license on a judgment; a community entrusts an organization with its most vulnerable members. AI can strengthen all of that — or, deployed carelessly, dissolve it in a single well-publicized failure.

The organizations that will lead behavioral health's AI era are not the ones that move first. They are the ones that can *prove* — not merely assert — that their AI is transparent, monitored, fair, data-protective, and human-governed. Responsible governance is how that proof is built. Which is why, in this field, it is not the constraint on the future. It is the foundation of it.

About PraxisOne

PraxisOne is the operating intelligence layer for behavioral healthcare. We connect the clinical, operational, and behavioral signals that already exist across an organization's environment, identify what requires attention, and turn intelligence into accountable, auditable action — with human oversight at every step, and governance designed in rather than bolted on. Our platform, Haleux, is delivered through Haleux Care, which brings governance to the front line of patient care and documentation, and Haleux Assure, which operationalizes QAPI as a continuous, evidence-backed discipline.

To discuss how governed-by-design operating intelligence could work in your environment, visit **praxisoneinc.com** or request a demonstration.

References

1. National Institute of Standards and Technology (NIST). AI Risk Management Framework (AI RMF 1.0). January 2023.
2. U.S. Food and Drug Administration. Clinical Decision Support Software — Guidance for Industry and FDA Staff. Final guidance, September 2022 (revised 2026).
3. Office of the National Coordinator for Health IT / HHS. Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing (HTI-1) Final Rule. 2024; predictive decision-support requirements effective January 1, 2025.
4. Texas Responsible Artificial Intelligence Governance Act (TRAIGA), effective January 1, 2026.
5. The Joint Commission & Coalition for Health AI (CHAI). Guidance on the Responsible Use of AI in Healthcare. 2025.
6. 42 CFR Part 2 (Confidentiality of Substance Use Disorder Patient Records); HIPAA, 45 CFR Parts 160 and 164.

Important notice

This document is provided for general informational and educational purposes only. It is not legal, regulatory, compliance, medical, or financial advice, and should not be relied upon as a substitute for the advice of qualified counsel or other professionals. Descriptions of laws, regulations, and frameworks are high-level summaries that may change over time and vary by jurisdiction and application; readers should consult the primary sources and their own advisors for current requirements. PraxisOne products are designed to align with responsible-AI principles and to support — not replace — the decisions of licensed clinical and administrative personnel; organizations remain responsible for their own compliance, and PraxisOne does not guarantee any particular regulatory or clinical outcome.