A DR GPT™ PUBLICATION

The DR GPT™ Healthcare AI Playbook

A Physician's Guide for CEOs, Boards, and Hospital Leaders

Harvey Castro, MD, MBA

Board-Certified Emergency Physician · 5x TEDx Speaker
Chief AI Officer, Phantom Space

EDITION 1.0 · SEPTEMBER 2026 · HARVEYCASTROMD.COM

CONTENTS

What is inside

00

Introduction: The Meeting Where This Usually Goes Wrong

I have been in the room for a lot of these meetings.

A vendor presents. The slide has a number on it, usually an area under the curve, usually somewhere north of 0.85. Somebody at the table says the word safety. Somebody else says efficiency. The clinical leader in the corner has a question but decides it will sound obstructionist, so the question does not get asked.

Nine months later a nurse on the fifth floor silences the same alert for the ninetieth time that week.

Nothing in that sequence was anyone's fault exactly. Everyone acted reasonably given what they knew. The problem is structural: the people who understood the technology were not the people who understood the shift, and no one in the room was responsible for connecting the two.

This playbook is written to make that connection.

I am a board-certified emergency physician. I still practice. I have also written more than thirty books on artificial intelligence in medicine, advised a national ministry of health, sat on a state medical association's health IT committee representing more than 55,000 physicians, and served as Chief AI Officer at an aerospace company where an unvalidated system carries consequences of a different kind.

That combination is the whole reason this document exists. Most healthcare AI guidance is written either by technologists who have never worked a shift or by clinicians who have never read a contract. You need both, in one head, at the same table.

What this is. A practical guide for the people who authorize the money, sign the contract, and answer for the outcome. Short chapters. Named evidence. Checklists you can take into a meeting on Monday.

What this is not. A survey of the technology, a vendor comparison, or a forecast. Predictions age badly. Judgment does not.

Read it in an hour. Use one chapter this quarter.

01

Why Healthcare AI Needs Physician Leadership

The failure modes are clinical, and they are invisible from outside the building

Here is the case that should be taught in every health system boardroom.

The Epic Sepsis Model was deployed at hundreds of US hospitals. It arrived with the electronic health record, which meant that in many organizations no one formally decided to adopt it. It was simply there.

When researchers at Michigan Medicine finally ran an independent evaluation across 38,455 hospitalizations, the results were sobering. The model produced an area under the receiver operating characteristic curve of 0.63. It failed to identify 1,709 of 2,552 patients with sepsis, roughly 67 percent. It generated alerts on 18 percent of all hospitalized patients, creating what the authors described as a large burden of alert fatigue. According to PubMed, that study is Wong A, Otles E, Donnelly JP, et al., *JAMA Internal Medicine*, 2021.

Read the last line of that abstract carefully. The authors wrote that widespread adoption of the model despite its poor performance raised fundamental concerns about sepsis management on a national level.

Now consider who could have caught this earlier.

Not the data science team, which did not have access to the proprietary model's internals. Not the executive team, which saw a vendor's published figures. Not the compliance function, which had no framework for a tool that arrived bundled with software already under contract.

The people who knew first were the nurses and physicians who noticed that the alert was usually wrong. They adjusted, rationally, by paying less attention to it. That adjustment was invisible in every dashboard the organization maintained.

Three things only a practicing clinician can do

Name the problem in clinical terms. Most AI projects begin with a capability looking for a use. A clinician can state the actual problem with enough precision that it becomes measurable: discharge summaries take twenty minutes each and are frequently late, or overnight CT reads for abdominal pain average ninety minutes. A problem stated that way tells you what to buy and when to stop.

Say how your patients differ. You do not need statistical training to ask which population a model was trained on and then describe how yours differs. Payer mix. Primary language. Comorbidity burden. How late in an illness people typically present. A physician can list those differences in thirty seconds. A vendor cannot list them for you.

Say no in public. Someone has to be able to stand in a leadership meeting, present evidence that a tool is not working, and recommend removing it. That requires standing, and it requires a culture where retiring a tool is a normal outcome rather than an admission of failure.

Leadership is not the same as consultation

Most health systems believe they already have physician input on AI. They have a chief medical information officer, or a clinician on the steering committee, or a medical director who reviews things.

Consultation and leadership are different, and the difference shows up under pressure.

Consultation means a clinician is asked for an opinion after the direction is set. The tool has been selected, the budget approved, the go-live scheduled. The clinician's role is to help with adoption. In that position, raising a fundamental objection means blowing up a project that six other people have already committed to, so most reasonable people do not raise it.

Leadership means a clinician defines the problem before the market is scanned, sets the validation requirement before the contract is negotiated, and holds the authority to stop the deployment without needing anyone's permission.

The test is simple. Ask what happens if your clinical AI lead says no next week. If the answer involves escalation, negotiation, or a conversation about being a team player, you have consultation. If the answer is that the deployment stops and the organization reconsiders, you have leadership.

Only one of those two protects patients, and only one of them satisfies what the 2027 statutes are reaching for.

The regulators have noticed the gap

State legislatures have started writing clinicians back into the loop by statute, because the market did not do it voluntarily.

Utah's SB 319 and Georgia's SB 544, both effective January 1, 2027, require that a licensed healthcare professional make an adverse determination independently rather than at a system's direction. Indiana's HB 1271, effective July 1, 2026, bars insurers from using AI as the sole basis to downgrade a claim without professional review. Washington's SB 5395 prohibits sole reliance on AI to deny or delay services. Delaware's HB 191 bars AI systems from professional licensure or from using protected titles.

There is a pattern in those bills. Legislators are legislating the presence of a clinician.

An organization that already puts a physician at the center of its AI decisions will find compliance mostly routine. An organization that has not will spend 2027 retrofitting.

> **ACTION BOX, CHAPTER ONE**
>
> Identify by name the practicing clinician accountable for AI oversight in your organization. If you cannot name one in ten seconds, that is the first appointment to make. Give the role protected time, not a title.

02

What CEOs Get Wrong About AI Adoption

Mistake one: treating the vendor's number as your number

Every clinical AI product arrives with a performance figure. That figure describes how the model behaved on the population it was built and tested on. It is not a promise about yours.

My patients in a Dallas emergency department differ from an academic medical center's cohort in language, insurance status, comorbidity burden, and how far into an illness they present. A model tuned on one distribution and dropped into another does not fail dramatically. It degrades quietly.

What to do instead. Require a local validation period before go-live, with a defined sample, defined metrics, a named clinical owner, and a stopping rule written in advance. Put the cost of that validation in the contract rather than discovering it later.

Mistake two: buying the model instead of the workflow

A perfectly calibrated model delivered into the wrong moment of a clinical day is still a bad product.

An overnight risk score that lands in a chart nobody opens until morning rounds has no clinical value regardless of its accuracy. A prompt requiring a nurse to leave the medication administration screen, read a paragraph, and click twice will be dismissed, because she is holding a syringe.

Compare that with ambient documentation, which works in most health systems for one reason: it sits inside a task clinicians already perform and removes work rather than adding it. In a multicenter study of 263 ambulatory clinicians across six health systems, burnout fell from 51.9 percent to 38.8 percent after thirty days, after-hours documentation dropped by a mean of 0.90 hours per day, and note-related cognitive task load improved by 2.64 points on a ten-point scale. According to PubMed, that is Olson KD, Meeker D, Troup M, et al., *JAMA Network Open*, 2025. A Stanford pilot with 48 physicians found the same direction of effect, per Shah SJ, Devon-Sand A, Ma SP, et al., *JAMIA*, 2025.

Same category of technology as the sepsis model. Opposite result. The difference is workflow design, and workflow design is a clinical act rather than a technical one.

Mistake three: assuming human oversight exists because a human is present

Human oversight only counts if disagreeing is cheap.

Three questions settle it. Can a clinician override the model without filing an incident report? Does overriding cost time she does not have? Does any dashboard, quality metric, or compensation formula penalize her for it?

If any answer goes the wrong way, the organization has automation with a witness.

Watch the override rate as a vital sign. A rate falling toward zero rarely means the model improved. It usually means people stopped checking.

Mistake four: measuring adoption instead of outcome

Adoption metrics are comfortable and nearly meaningless. Logins, percentage of notes generated, seats deployed. None of them tell you whether anything got better.

Four numbers per clinical tool, reviewed quarterly, tell you what you actually own:

1. Local performance on current patients, not the vendor's original cohort
2. Alerts per 100 admissions
3. Override rate and its direction over time

4. Performance stratified by language, payer, race, and site

If a tool cannot produce those four, the organization does not know what it has running.

Mistake five: letting the savings evaporate

This is the one that costs organizations their next initiative.

A system deploys ambient documentation, clinicians get roughly an hour back, and finance responds by increasing panel size until the hour is gone. Burnout returns. Trust in the tool collapses. The next project meets a much harder audience.

Decide in advance what the recovered time is for: patient access, teaching, a shorter workday, or a smaller locums line. Write it down before go-live, because afterward it becomes a budget negotiation and clinicians lose that one every time.

> **ACTION BOX, CHAPTER TWO**
>
> Pick your largest AI deployment. Ask for those four numbers in writing within thirty days. The speed and completeness of the answer tells you more about your governance than any policy document.

03

The DR GPT™ Framework for Safe AI in Hospitals

Stop inventing. Credential the model like a clinician.

Every hospital already knows how to govern something that makes clinical decisions, carries real risk, performs differently at different institutions, and occasionally has to be stopped.

It is called a doctor.

You credential them. You grant privileges within a defined scope. You supervise until competence is demonstrated. You review performance on a schedule. When performance falls apart, you act.

That machinery has existed for a century and it works. Most AI governance efforts fail because someone tried to build a parallel structure from scratch, staffed it with a committee meeting quarterly, and produced a policy nobody in the building has read.

The framework below maps AI oversight onto the credentialing process your medical staff office already runs.

Stage one: Credential

No model touches a patient until four questions are answered in writing.

Where did it come from, and who built it. Include whether it arrived bundled with existing software, which is how a surprising number of tools enter a hospital without a decision.

What population was it trained on. Then state, explicitly, how yours differs.

How does it perform on your patients. Measured locally, before go-live, on a defined sample.

Who signed off, by name. Not a department. A person, ideally a practicing clinician, who still works here.

That last requirement is the one organizations skip, and it is the one that makes everything downstream possible. Accountability without a name is not accountability.

Stage two: Privilege

A credentialed clinician does not get to do everything. A cardiologist does not perform craniotomies.

Write the scope in plain language. Which patients. Which settings. Which decisions the tool may inform, and which decisions are explicitly out of bounds. Documentation tools carry different privileges than triage models, and triage models carry different privileges than anything touching a treatment recommendation.

Then name the supervising role. Not a committee. A role that can be paged.

Stage three: Supervise

Human oversight is a design property, not a policy statement.

Make the override path shorter than the agreement path for any decision with real downside. Show the drivers behind an output rather than only the number. Flag when the model is operating outside its validated range. Then check whether anything in the organization punishes a clinician for disagreeing.

Stage four: Evaluate

Your medical staff bylaws already require periodic performance review. Apply the same discipline on a fixed cadence.

Discrimination and calibration on current patients. Alert volume and override rate. Performance stratified across your populations. Drift measured against the baseline set at credentialing.

Set the action threshold before you need it. A number written in advance is a governance decision. The same number debated after an incident is a negotiation.

Stage five: Peer review and revoke

Two things must exist.

A path for a frontline clinician to report that a model harmed or nearly harmed a patient, routed to the same committee that handles other safety events. And a documented authority to shut a model off, with a named person who holds it.

Most AI policies have no off switch, no named holder, and no precedent for using one.

Retire one tool publicly in the first year. The organization learns more from that single act than from any binder.

The framework applied: a worked example

Suppose your organization is evaluating a deterioration-prediction model for medical-surgical floors.

Credential. The vendor reports an area under the curve of 0.87, developed on 400,000 admissions across twelve academic centers. Your patient population is 38 percent Medicaid, 22 percent primarily Spanish-speaking, and includes two community hospitals not represented in that development set. You require a 90-day silent evaluation on your own data before any alert reaches a clinician, with the metric agreed in advance and a named hospitalist as approver.

Privilege. The tool may flag adult inpatients on medical-surgical units. It may not be used in the emergency department, in obstetrics, or for pediatric patients. It informs a rapid response consideration. It does not trigger an automatic transfer.

Supervise. The alert opens with the three variables driving the score, and flags when the patient falls outside the validated range. Dismissal takes one click with an optional reason, and the nursing productivity dashboard is checked to confirm dismissals count against no one.

Evaluate. Monthly review of discrimination, alerts per 100 admissions, dismissal rate, and performance split by language, payer, and site. Pre-agreed threshold: if alert volume exceeds the agreed ceiling for two consecutive months, or subgroup performance falls below the floor, the tool moves to reduced scope pending review.

Peer review and revoke. Frontline reports route to the patient safety committee. The chief medical officer holds documented authority to suspend the tool, written into the vendor agreement so suspension is not a breach.

Notice how much is decided before go-live. One committee cycle, and the organization already knows what it will do when the tool underperforms.

How the framework maps to certification

On June 2, 2026, The Joint Commission launched a voluntary Responsible Use of AI in Healthcare certification covering five domains: governance, data management, risk and bias reduction, monitoring and validation of safety and effectiveness, and transparency with education and training. Any organization may apply, accredited or not.

An organization running the five stages above is already producing most of the evidence those domains require. Governance built on machinery your staff already understands survives turnover, audits, and the next technology cycle.

> **ACTION BOX, CHAPTER THREE**
>
> Take your three highest-risk AI tools through stage one this quarter. Expect the exercise to surface at least one tool nobody formally approved.

04

AI Governance Without Killing Innovation

The objection, stated fairly

The argument against governance goes like this. Healthcare moves slowly enough already. Every new control adds a committee, a form, and six weeks. Competitors who move faster will capture the benefit while you are still validating.

I take that seriously. Slow governance is a real cost, and healthcare has a long record of building processes that outlive their purpose.

But the evidence points the other way.

What ungoverned adoption actually costs

The buying has already happened. Among the 2,784 US hospitals running Epic, 62.6 percent had adopted ambient AI documentation tools as of June 2025, per Yang F and Graetz I in the *American Journal of Managed Care*, 2026.

The oversight has not kept pace. Black Book Research surveyed 182 hospital leaders between October and November 2025 and found that only 29 percent had implemented and enforced policies covering AI model inventory, lineage, and sign-offs, with 48 percent still developing them. Only 22 percent were highly confident they could produce a complete AI audit trail within 30 days for a regulator or payer, falling to 15 percent among small hospitals.

The reported obstacles are instructive. Forty-one percent named insufficient vendor documentation. Thirty-three percent named unclear ownership across IT, quality, safety, and compliance.

Neither of those is a governance-is-too-slow problem. Both are governance-never-happened problems, and both are expensive to fix after the fact: a contract renegotiated from a weak position, a year spent unwinding alert fatigue, a board meeting where nobody can produce a list of what is running.

The budget question nobody asks

The median share of 2026 IT and quality or safety budgets allocated to AI governance and safety is 4.2 percent. Large systems allocate 6.8 percent. Smaller hospitals allocate 2.3 percent. Only 26 percent of surveyed organizations plan to increase that by two or more percentage points in 2026, and 18 percent plan no increase at all.

If your AI portfolio is growing and your oversight budget is flat, the board is carrying an unfunded control. That is a finding, not an opinion.

Four principles that keep governance fast

Tier by risk, not by novelty. A scheduling optimizer and a sepsis predictor should not travel the same approval path. Define three tiers with different requirements and let low-risk tools move quickly.

Set decision deadlines for the committee, not only for the requester. A governance body that can block indefinitely will be routed around. Give it a clock.

Prefer standing rules over case-by-case review. Write the local validation requirement, the monitoring cadence, and the contract clauses once. Most requests then become checklist items rather than debates.

Make retirement routine. Governance that only ever adds tools is procurement wearing a different badge.

The three tiers, written out

Rather than leaving this as an exercise, here is a starting structure. Adapt the thresholds to your organization, but keep it to one page.

Tier 1. Administrative, no patient-specific output. Scheduling optimization, supply forecasting, transcription of non-clinical meetings, staffing models. *Requirements:* privacy and security review, named business owner, annual re-review. *Decision clock:* two weeks.

Tier 2. Patient-specific but not decision-directing. Ambient documentation, chart summarization, patient messaging drafts, coding assistance, translation support. *Requirements:* everything in tier 1, plus a clinician approver, a defined scope of use, a documented review step before anything clinically consequential is finalized, and quarterly monitoring of usage and error reports. *Decision clock:* four to six weeks.

Tier 3. Informs or influences a clinical or payment decision. Deterioration and sepsis prediction, triage and acuity scoring, imaging interpretation support, risk stratification driving outreach, anything touching utilization review or an adverse determination. *Requirements:* full credentialing per chapter three, local validation before go-live, a written stopping rule, monthly monitoring with stratified performance, a named supervising role, a documented off-switch authority, and board-level visibility. *Decision clock:* one quarter, with a silent evaluation period that does not count against it.

Two rules make the tiering hold. First, a tool bundled inside software you already own still gets tiered, because bundling is how ungoverned tools enter a hospital. Second, tier is assigned by what the output influences, not by how the vendor describes it. A product marketed as documentation support that surfaces a risk score is tier 3.

Most of what an organization runs will land in tiers 1 and 2, which is the point. The heavy process applies to the small number of tools where it earns its cost.

Why clinicians adopt governed tools faster

This is the part executives underestimate.

Clinicians use tools they trust, and trust comes from knowing the thing was checked, that someone competent is watching it, and that saying no carries no penalty. A tool introduced

with local validation data, an honest account of its limits, and a named owner gets adopted faster than one introduced with a vendor slide and an enthusiasm memo.

Governance is not the brake. In healthcare it is the accelerator, because it is the only thing that converts a deployment into sustained clinical use.

> **ACTION BOX, CHAPTER FOUR**
>
> Write your risk tiers this month. Three tiers, one page, with the requirements and the decision clock for each. Nothing improves governance speed faster than removing the question of how much process applies.

05

How to Prepare Clinicians for AI

Start by admitting what the last technology cycle did to them

Any conversation about preparing clinicians for AI has to begin with an honest acknowledgment. The last major technology transformation in medicine, the move to electronic health records, was sold as a tool to help clinicians and delivered, for many of them, a documentation burden that has not lifted.

Clinicians remember. When leadership announces that a new system will give them time back, a portion of the room is doing arithmetic on the last promise.

You do not overcome that with enthusiasm. You overcome it by being specific, measuring honestly, and letting the first deployment be one that unambiguously removes work.

Sequence matters more than curriculum

Most organizations start AI readiness with education. I would start with a win.

Deploy something that takes work away before you deploy anything that adds a decision. Ambient documentation is the obvious candidate, and the evidence supports it. Burnout dropping from 51.9 percent to 38.8 percent in thirty days, with roughly 0.90 fewer hours per day of after-hours documentation, changes how the next announcement is received.

Then educate, with credibility you have earned rather than borrowed.

What clinicians actually need to know

Not model architecture. Four things.

How this specific tool performs here. Local numbers, including the unflattering ones. A tool presented with its limits is trusted more than one presented as flawless.

What it is not allowed to do. Scope stated plainly reduces both overuse and reflexive rejection.

How to disagree. Where the override lives, how long it takes, and confirmation that using it carries no penalty.

How to report a problem. The same channel used for other safety events, with the expectation of a response.

That is a thirty-minute session, not a semester.

A thirty-minute training template

Use this structure for any clinical AI rollout. It fits in an existing staff meeting.

Minutes 0 to 5. The problem. State the clinical problem this tool addresses, in the words the unit would use. Not the technology. The problem.

Minutes 5 to 12. How it performs here. Show the local validation result, including what it does poorly. Name the population it was developed on and how yours differs. This is the segment that earns the room.

Minutes 12 to 18. What it may and may not do. The privilege scope, stated plainly, with two or three concrete examples of decisions that remain entirely clinical.

Minutes 18 to 24. How to disagree, and automation bias. Demonstrate the override live. State that no metric penalizes its use. Then describe automation bias by name and give one example. Clinicians recognize it instantly, and recognition is most of the defense.

Minutes 24 to 30. How to report a problem, and what happens next. Name the channel, the committee, and the response expectation. State the monitoring cadence and the

threshold that would take the tool out of service.

Close by naming the person accountable and how to reach them. A room that leaves knowing a name behaves differently than a room that leaves knowing a percentage.

Automation bias deserves its own conversation

The failure that worries me most is the one that looks like success.

A confident output on a screen shifts a clinician's threshold, particularly when that clinician is tired, junior, or covering an unfamiliar service. The model says low risk, the patient goes home, and the chart records a clinician who agreed with a computer at four in the morning.

Name this explicitly in training. Clinicians recognize it immediately once it is described, and recognition is most of the defense. Pair it with design: show uncertainty, show the drivers behind the output, and flag when the model is operating outside its validated range.

Give the skeptics a real job

The physician who objects loudest in the meeting is often your most valuable implementation asset, and organizations routinely waste them.

Put the skeptics on the validation review. Ask them to define the stopping rule. Let them design the override path. A skeptic given authority becomes a rigorous evaluator. A skeptic dismissed becomes a permanent obstacle, and they talk to colleagues.

Do not forget everyone who is not a physician

Nurses, pharmacists, therapists, technicians, and schedulers usually encounter these tools more often than physicians do, and are usually consulted less.

The nursing workforce picture makes this urgent rather than merely polite. HRSA projects a shortfall of 108,960 full-time registered nurses by 2038, with nonmetropolitan areas

facing an 11 percent shortage against 2 percent in metro areas, and a projected shortfall of 245,950 licensed practical nurses with only 70 percent of demand met.

Any AI initiative that adds clicks to a nursing workflow is subtracting from a workforce you cannot replace.

> **ACTION BOX, CHAPTER FIVE**
>
> Before your next deployment, sit through three shifts where the tool will be used. Count clicks and interruptions. Bring numbers back to the design conversation. Nothing you learn in a conference room will match one night on the floor.

06

The Future of Medicine Is Human Plus AI

The framing that keeps getting this wrong

The question was never whether a machine can do a clinician's job. The question is how much of a clinician's day is not clinical work at all, and how much of that can be given back.

Start with arithmetic, because arithmetic is what makes this urgent.

The Association of American Medical Colleges projects a shortage of up to 86,000 physicians by 2036. The drivers are demographic and largely fixed: the population over 65 grows 34.1 percent, the population over 75 grows 54.7 percent, and 20 percent of the current clinical physician workforce is already 65 or older with another 22 percent between 55 and 64.

Training pipelines take a decade. Compensation competition moves clinicians between systems without creating new ones. The supply side is mostly determined.

What remains adjustable is the work itself.

What the best version of this looks like

The clearest published example comes from breast cancer screening in Sweden.

The MASAI trial randomized 105,934 women in the Swedish national screening program to AI-supported screening or standard double reading. AI-supported screening detected 6.4 cancers per 1,000 participants against 5.0 per 1,000 in the control group, a 29 percent

increase, with the additional cancers mainly small, lymph-node negative invasive cancers. False positives did not rise significantly. Screen-reading workload fell 44.2 percent. According to PubMed, that is Hernström V, Josefsson V, Sartor H, et al., *Lancet Digital Health*, 2025.

Study the design rather than the headline. The AI triaged which examinations needed one reader versus two. The radiologist still read and still decided. More cancers found, fewer reading hours consumed, no radiologist replaced.

That is the shape of a healthcare AI investment worth funding.

The larger prize is healthspan, not lifespan

Medicine got very good at extending life. It has been much worse at improving the end of it.

A study across all 183 World Health Organization member states found the global gap between lifespan and healthspan has widened to 9.6 years. The United States has the largest gap on earth at 12.4 years, driven by the burden of noncommunicable disease. According to PubMed, that is Garmany A and Terzic A, *JAMA Network Open*, 2024.

The average American spends more than a decade at the end of life in an impaired health state. No algorithm fixes that alone. But earlier detection, outreach to patients who never schedule the appointment, and clinicians with enough minutes to actually do prevention all move the number in the right direction.

Where AI must not go

The credibility of everything above depends on being equally direct about the boundary.

Adverse determinations belong to licensed humans, and multiple states now require it. Behavioral health has drawn the hardest lines: Maine's HB 2082 restricts mental health professionals to administrative uses of AI, Arizona requires documented informed consent before AI is involved in behavioral health services beginning January 1, 2027, and Idaho

and Nebraska passed Conversational AI Safety Acts effective July 1, 2027 requiring chatbots to disclose that they are AI and to implement suicide-response protocols.

Those boundaries are correct, for a reason that has nothing to do with technology. A person in crisis needs a person.

The equity problem hiding inside the good news

The systems that most need capacity relief are getting these tools last.

Ambient AI adoption ran at 64.7 percent in metropolitan hospitals against 54.3 percent outside them, 70.2 percent at nonprofits against 28.8 percent at for-profit facilities, and 67.6 percent among hospitals in the highest operating-margin quartile against 58.0 percent in the lowest.

Now overlay HRSA's projection that nonmetropolitan registered nurse shortages reach 11 percent by 2038 against 2 percent in metro areas. The places with the thinnest workforce are adopting the capacity tools slowest.

That gap closes through purchasing coalitions, regional collaboratives, and health systems that extend enterprise agreements to rural affiliates rather than piloting only at the flagship. It does not close on its own.

> **ACTION BOX, CHAPTER SIX**
>
> Name the specific clinical bottleneck in your organization where a scarce specialist's reading or review time is the constraint. That is where a MASAI-style intervention belongs, and it is a better first investment than a diagnostic model with no capacity effect.

07

Takeaways for Boards and Leadership Teams

Ten questions, and what a good answer sounds like

Directors do not need to understand model architecture. They need to ask questions management cannot answer with a slide.

1. How many AI tools are running here today, and who maintains that list? *Good:* a number, an owner, and a refresh date. *Bad:* a number with no owner.

2. Which touch a clinical decision, and which touch a payment decision? The risk profiles differ, and so does the law. If you operate across state lines, ask how policy handles the variation.

3. What did local validation show before each clinical tool went live? *Good:* performance on your own patients, with a date. *Bad:* the vendor's published figure.

4. Who signed the approval, by name, and do they still work here? Turnover quietly orphans approvals.

5. What is our override rate, and which direction is it moving? Ask whether any metric or compensation formula penalizes a clinician for overriding.

6. Have we measured performance across our patient populations? Stratified by language, payer, race, and site. A model that works on one population and fails on another is a quality failure before it is a legal one.

7. What is our process for turning a model off, and have we ever used it? *Good:* a named authority, a defined trigger, and at least one tool retired.

8. If a regulator requested a complete AI audit trail tomorrow, how many days? Ask for a number in days. Ask whether the organization intends to pursue Joint Commission certification, and if not, why not.

9. What percentage of the IT and quality budget funds AI oversight? Median is 4.2 percent. Below that with a growing portfolio is an unfunded control.

10. What is our exposure if a vendor changes a model without telling us? Silent model updates are a new device in your hospital that nobody credentialed.

The two mistakes boards make

Delegating the whole subject to the technology committee. Clinical AI creates quality, workforce, liability, and reputational risk. Routing it to IT governance alone means the people who understand patient safety never see it. Put it on the quality committee with IT and legal present.

Treating one education session as oversight. A guest speaker, a deck, and a nodding board is not a control. I say that as someone who is frequently the guest speaker. Only a standing agenda item with numbers attached changes an organization.

What belongs in the board minutes

Oversight that leaves no record is difficult to demonstrate later, whether to a regulator, an insurer, or a plaintiff's counsel.

At minimum, the minutes should reflect that the board received a current AI inventory with named owners, reviewed performance metrics for tools in the highest risk tier, was informed of any tool added or retired during the period, and was told the organization's current audit-trail readiness in days.

That is four sentences per quarter. It is also the difference between an organization that governed and an organization that intended to.

The first ninety days

If your board or executive team is starting from close to zero, this is the sequence I would run.

Days 1 to 30. Inventory. Ask in writing what AI is running and who approved each tool. Expect an incomplete answer, and treat the incompleteness as the finding rather than an embarrassment. Appoint a named clinical owner with protected time.

Days 31 to 60. Triage. Sort the inventory into three risk tiers. Take the top tier through credentialing: provenance, training population, local performance, named approver. Pull one contract and check it for local validation, monitoring obligations, update notification, and data portability.

Days 61 to 90. Decide something. Retune a threshold, add a monitoring metric, or retire a tool nobody uses. Report the decision to the board. Put AI oversight on the quality committee's standing agenda with the four operating metrics attached.

Ninety days will not make you compliant with everything arriving in 2027. It will make you an organization that knows what it owns, which is more than roughly seven in ten currently can say.

The one thing to remember

AI will not fail your hospital.

Decisions made about AI without a physician in the room are what fail hospitals. Everything in this playbook is a variation on that sentence.

APPENDIX

Appendix A. The One-Page AI Inventory

For each tool, one row. If a cell cannot be filled, that is the work.

FIELD	WHAT TO CAPTURE
Tool name and vendor	Include tools bundled with existing software
Function	Documentation, prediction, triage, imaging, administrative, payment
Risk tier	1 low, 2 moderate, 3 clinical decision or payment determination
Training population	As stated by the vendor
Local validation	Metric, sample, date, result
Named approver	Person and role, plus whether still employed
Supervising role	Who gets paged
Monitoring cadence	And the threshold that triggers action
Alerts per 100 admissions	Current and trend
Override rate	Current and trend
Stratified performance	By language, payer, race, site
Contract review date	And renewal date
Off-switch authority	Named person

Appendix B. Contract Clauses to Require

Local validation before go-live. Defined sample, defined metrics, named clinical owner, and cost allocation stated in the agreement.

Ongoing performance monitoring. Vendor obligation to report performance against an agreed baseline on an agreed cadence.

Model change notification. Written notice before any retraining or model update reaches production, with the right to re-validate.

Documentation. A model card or equivalent that a quality committee can actually read. Note that 41 percent of surveyed hospital leaders named insufficient vendor documentation as their primary obstacle to audit readiness, which makes this a negotiation item rather than an afterthought.

Audit trail access. The organization can produce a complete record of model inputs, outputs, versions, and approvals on demand.

Stopping rule. A defined performance threshold that permits suspension without penalty.

Data portability on exit. Model outputs, audit logs, and underlying data leave with you.

A small hospital has the most bargaining power at signature and almost none afterward. Negotiate these before the purchase order, or accept that you will not get them.

Appendix C. Objections You Will Hear, and Honest Answers

"Our vendor already validated it." On their population. Ask which one, then say out loud how yours differs. Validation is not a certificate the vendor holds. It is a measurement you take.

"We do not have the data science capacity for local validation." Most local validation is a retrospective comparison against outcomes you already record, not original research. Where capacity is genuinely absent, join a regional collaborative or purchasing coalition, and write the requirement into the contract so the cost sits with the vendor.

"Governance will slow us down and competitors will get ahead." The organizations spending 2026 unwinding alert fatigue and renegotiating contracts from a weak position are the ones that skipped it. Tier by risk so low-risk tools move quickly, and give the committee a decision clock.

"Our clinicians will reject anything we deploy." Clinicians reject tools that add work and tools they do not trust. Deploy something that removes work first, present it with its limits, and give people a real override. Adoption follows.

"This is an IT matter." Clinical AI creates quality, workforce, liability, and reputational risk. Technology governance alone means the people who understand patient safety never see it.

"We are too small for this." Small hospitals carry more of this risk, not less. Only 15 percent of small hospitals surveyed were confident they could produce an audit trail within 30 days. The framework in chapter three requires a named physician, a spreadsheet, and a standing agenda item.

"The regulation is still unsettled, so let us wait." Effective dates have already passed in Indiana and Maryland, with more arriving through 2027, and a certification program

launched in June 2026. Waiting is now a decision with a date attached to it.

"We already have an AI committee." Ask when it last declined something. A body that has never said no is a routing function, not a control.

"Nobody else is doing this." Twenty-nine percent have enforced inventory and sign-off policies. Being early here is inexpensive and being late is not.

"What if we validate and the tool fails?" Then you learned it for the cost of a validation period rather than the cost of an adverse event, a retraction, and a year of lost clinical trust. That is the entire return on this work.

About the Author

Harvey Castro, MD, MBA is a board-certified emergency physician, 5x TEDx speaker, and author of more than thirty books on artificial intelligence and healthcare, including *AI in Emergency Medicine* (Wiley) and *AI and Healthcare, 2nd Edition* (Routledge).

He serves on Singapore's Ministry of Health Regulatory Advisory Panel, advises the Texas Medical Association's Committee on Health Information Technology representing more than 55,000 Texas physicians, teaches in the UTSA and UT Health Medicine and AI dual-degree program, and serves as Chief AI Officer at Phantom Space Corporation. He sits on the editorial boards of *AI in Precision Oncology* and *The American Journal of Healthcare Strategy*, and has delivered keynotes at HIMSS Europe and HIMSS AI and across China, Australia, and Germany. D Magazine named him ER Doctor of the Year from 2014 through 2022. He is a native and bilingual Spanish speaker.

He still works clinical shifts.

REFERENCES

Sources

Wong A, Otles E, Donnelly JP, et al. External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Internal Medicine*. 2021;181(8):1065-1070. doi:10.1001/jamainternmed.2021.2626

Olson KD, Meeker D, Troup M, et al. Use of Ambient AI Scribes to Reduce Administrative Burden and Professional Burnout. *JAMA Network Open*. 2025;8(10):e2534976. doi:10.1001/jamanetworkopen.2025.34976

Shah SJ, Devon-Sand A, Ma SP, et al. Ambient artificial intelligence scribes: physician burnout and perspectives on usability and documentation burden. *Journal of the American Medical Informatics Association*. 2025;32(2):375-380. doi:10.1093/jamia/ocae295

Hernström V, Josefsson V, Sartor H, et al. Screening performance and characteristics of breast cancer detected in the Mammography Screening with Artificial Intelligence trial (MASAI). *Lancet Digital Health*. 2025;7(3):e175-e183. doi:10.1016/S2589-7500(24)00267-X

Garmany A, Terzic A. Global Healthspan-Lifespan Gaps Among 183 World Health Organization Member States. *JAMA Network Open*. 2024;7(12):e2450241. doi:10.1001/jamanetworkopen.2024.50241

Yang F, Graetz I. Ambient AI Tool Adoption in US Hospitals and Associated Factors. *American Journal of Managed Care*. 2026;32(1):e25-e30.

Association of American Medical Colleges. The Complexities of Physician Supply and Demand: Projections From 2021 to 2036. March 2024.

HRSA Bureau of Health Workforce. Nurse Workforce Projections, 2023-2038.

Black Book Research. 2026 Health System AI Governance Resource Guide, surveyed October 15 to November 8, 2025, as reported by Becker's Hospital Review.

The Joint Commission. Responsible Use of AI in Healthcare certification, announced June 2, 2026. Joint Commission and Coalition for Health AI, Responsible Use of AI in Healthcare guidance, September 2025.

Holland & Knight. States Continue Efforts to Regulate AI in Healthcare: A Review of Legislation Passed in 2026. May 2026.

Clinical study findings cited above were retrieved from PubMed.

BRING THIS CONVERSATION TO YOUR BOARD

Book Harvey Castro, MD DR GPT™

Board advisory · Executive and leadership retreats
Conference keynotes, in English or Spanish

---

www.HarveyCastroMD.com

linkedin.com/in/harveycastromd

#DRGPT

ALSO REPRESENTED BY EXECUTIVE SPEAKERS BUREAU AND BIGSPEAK