

Regulation, Governance, and the Future

Study Guide

by Richard J. Young, Ph.D.

richard@deepneuro.ai

Chapter 25 of "Healthcare Analytics and AI: Building Systems That Actually Work"

The essentials: the checkpoint between a working model and the bedside.

One-sentence thesis. A technically excellent model reaches a patient only if it clears a regulatory pathway, is held up by organizational governance, and answers the liability question before it fails, because most healthcare AI failures are governance failures, not model failures.

PART 1: FDA SaMD and the Moving-Target Problem

Software as a Medical Device: Three Pathways

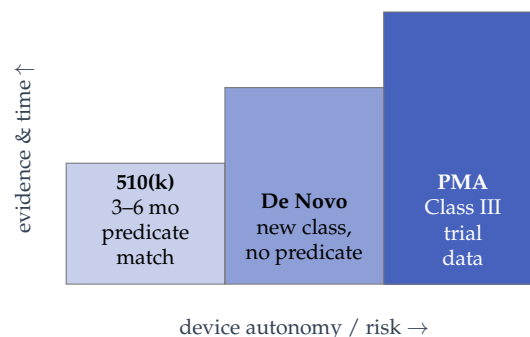
The gate opens with a definition: if your algorithm diagnoses, treats, or prevents disease, the FDA calls it a device, and there are already more than **1,300** FDA-authorized AI devices (258 cleared in 2025 alone, roughly 75 to 80 percent of them in radiology). Three doors lead into the clinic, each demanding a different amount of evidence to unlock. **510(k)** clears a device as substantially equivalent to an existing predicate (3 to 6 months, no clinical trial, most radiology tools). **De Novo** creates a new classification when no predicate exists (IDx-DR, the first autonomous diagnostic,

2018). **PMA** is for the highest-risk Class III devices and requires trial data.

ANALOGY

Three doors into the clinic, each needing a different amount of evidence to unlock.

Watch out. The 510(k) trap is that a device can be "substantially equivalent" to a mediocre predicate and enter the market without ever showing it improves care. As AI moves toward autonomous decisions, more tools will need PMA, and the infrastructure was not built to absorb that.



The three FDA device pathways. Evidence and time rise with device risk: most radiology tools take the 510(k) door, IDx-DR opened the De Novo door, and the highest-risk Class III devices need a full PMA trial.

Predetermined Change Control Plans (PCCP)

Traditional regulation assumes a static product, but ML models learn and drift, so how do you approve a moving target? The PCCP is the answer: filed with the initial submission (finalized December 2024), it states in advance which post-market changes (retraining on same-population data, sensitivity tuning within bounds) are pre-approved. It works like a renova-

tion permit that tells regulators which walls you may move without re-approving the whole building. In-scope changes ship without a new submission; out-of-scope changes still require review.

The PCCP turns continuous monitoring into a compliance mechanism, but it forces manufacturers to forecast every future change at submission time, creating pressure to write it broadly, which the FDA pushes

back against. That tension will define AI device regulation for a decade.

Watch out. Agentic systems that improvise in novel

situations operate outside any predetermined scope by design. The gap between adaptive AI and the PCCP assumption is widening.

Four Risk Tiers and Extraterritorial Reach

Where the FDA inspects each car model, the EU writes binding traffic laws for the whole road, and fines you a share of global revenue for breaking them (up to 35 million euros or **7 percent** of global turnover). Regulation (EU) 2024/1689 sorts AI into unacceptable (banned), high-risk (mandatory requirements), limited risk (transparency), and minimal risk (none). Clinical AI is automatically high-risk, triggering mandatory lifecycle risk management, data governance (relevant, representative, bias-audited data), audit logging, transparency, and meaningful human oversight. Core obligations apply from August 2026.

Article 2 gives it extraterritorial reach: if your system touches a patient in any EU state, you comply regardless of where you are headquartered. Like GDPR, its requirements are becoming the de facto global floor, so US teams building for global deployment simply satisfy both regimes. The philosophical split is the key to remembering it: the FDA asks “does this device meet safety and effectiveness for clearance?” while the EU asks “does the entire lifecycle satisfy mandatory risk, data, transparency, and oversight requirements?” The broader EU bar sets the compliance floor.

State Sandboxes and the Practice-of-Medicine Gap

The FDA regulates devices; states regulate the practice of medicine, and between them sits a gap for AI

that makes clinical decisions without being a device. Utah answered first: in January 2026 it let Doctronic autonomously renew prescriptions inside a 12-month regulatory sandbox, a fenced test track rather than a blank check, with temporary relaxed rules, heavy monitoring, and the track closing if it fails. The reporting requirement then earned its keep. The first public findings, released in May 2026 and covering roughly four months, show the AI recommending renewal in **72 percent** of the cases it triaged and escalating the rest, with reviewing physicians agreeing with those renewals 91 percent of the time (97 percent counting either of two reviewers) and judging 69 percent of the escalations warranted. The report withheld total case counts, describing volume only as limited, and Doctronic must sustain concordance above 98 percent before its autonomy expands, so the honest reading is a system that triages conservatively, agrees with physicians most of the time, and has not yet run at a scale that could establish safety. Texas (TRAIGA), Arizona, and others followed; by early 2026, 47 states had introduced over **250** AI-in-healthcare bills.

Watch out. The unresolved question is whether an AI that renews prescriptions is practicing medicine (regulate under medical-practice acts), a device (FDA), or consumer software (no one). The current answer, a state commerce department running a sandbox, satisfies neither the AMA nor the technology industry, nor the patient whose metformin it renews.

Model Cards: The Nutrition Label

Before any system reaches a patient, someone should be able to read what it does and how well it does it. A model card is that disclosure: model identity and intended use, training-data demographics, performance stratified by subgroup, known failure modes, deployment and integration details, and regulatory status. You would not eat a food that refused to list its ingredients, and the HTI-1 rule (January 2025) made this kind of transparency a requirement for clinical decision support. The reason the subgroup line matters is arithmetic: a model that is 92 percent accurate overall but 74 percent for Black patients is not a 92 percent model, it has a racial performance gap the aggregate conceals (Chapter 23). If you cannot complete a model card and a deployment checklist, you are not ready to deploy.

The PPTO Governance Framework

Governance is the air-traffic control tower; without it, technically capable systems still collide. The PPTO framework names its four parts. **People:** a multidisciplinary council whose authority sits high enough (board-level) to pause or retire an unsafe algorithm independent of the unit that deployed it. **Process:** three lines of defense (operations, compliance, independent audit) with shadow-testing gates. **Technology:** model registries, drift dashboards, audit logs, automated alerting. **Operations:** risk-tiered controls and rollback drills, so a scheduling optimizer gets an annual review while an autonomous insulin agent gets continuous monitoring and rollback drills.

KEY IDEA

Most failures were governance failures. Epic's sepsis model, the Optum algorithm, and IBM Watson for Oncology were never stopped by regulation; no structure had the authority, data, or mandate to shut them down.

Watch out. Shadow AI. A December 2025 survey found 17 percent of healthcare workers used unauthorized AI tools (57 percent had encountered it), adding an average of **\$670,000** to breach costs. The fix is responsive governance, not prohibition, because clinicians reach for shadow tools when the approved ones are too slow.

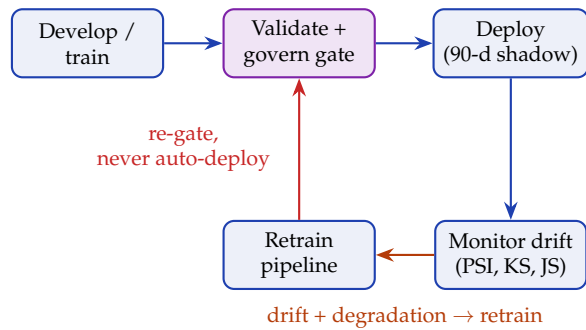
MLOps and Post-Market Monitoring

Fewer than **30 percent** of healthcare AI pilots reach production, and the blocker is operational readiness, not model performance. Production clinical AI needs containerized inference, shadow deployment (Mayo requires a 90-day shadow period), and automated drift detection (Population Stability Index, Kolmogorov-Smirnov, Jensen-Shannon) that fires when input distributions shift. This is pharmacovigilance for algorithms: drugs are watched for adverse events after approval, and models must be watched for drift after deployment. Drift plus degradation triggers a retraining pipeline, but never automatic deployment; the retrained model passes back through validation and governance gates.

The reason to watch is that a model at 0.92 AUC on Tuesday can hit 0.78 on Thursday when the case mix shifts (flu season, an EHR coding change, a pandemic). Bigger is not better either: a 340M-parameter model (BioLinkBERT) beat models up to ten times larger on a clinical task while keeping PHI on-premises, so architecture choice is a governance decision, not just a technical one.

Watch out. AI validating AI creates a recursive problem. If the testing tool has a blind spot, the certification is meaningless but looks rigorous. Who validates the validator?

Domain	Key question	Concrete artifact
People	Who is responsible?	Governance council charter
Process	What are the rules?	Shadow-testing protocol
Technology	What enforces it?	Model registry + drift dashboard
Operations	How does it sustain?	Tier 1/2/3 classification matrix



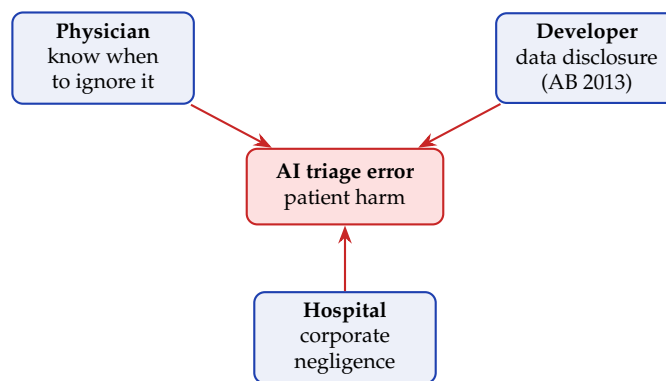
Algorithmovigilance as a loop. Detected drift triggers a retraining pipeline, but the retrained model always passes back through validation and governance, never straight to deployment.

Who Is Liable When AI Fails?

Start with the harm, not the doctrine: an AI triage system undertriaged a heart attack, the patient waited 47 minutes, and suffered avoidable myocardial damage. Who pays? Healthcare delegated the work to algorithms faster than it assigned responsibility for the consequences, and liability now splits three ways. **Physicians** bear ultimate responsibility and must know when to use AI and when to ignore it. **Developers** were historically shielded (“we do not practice medicine”), but AI malpractice claims rose 14

percent from 2022 to 2024 and California’s AB 2013 now forces training-data disclosure. **Hospitals** face corporate-negligence claims for deploying without vetting or training.

Watch out. A paradox is forming: physicians are liable if they follow a wrong AI recommendation, and may soon be liable if they fail to use a right one (missing a nodule an FDA-cleared tool would have caught). The only safe harbor is clinical judgment, which requires training and time, both in short supply.



Liability splits three ways. When an AI triage error harms a patient, responsibility distributes across the physician who used it, the developer who built it, and the hospital that deployed it.

The Trust Gap and the Forces Ahead

Trust is the last gate, and the numbers show a gap. In the 2025 Philips survey, 79 percent of professionals were optimistic about AI but only 59 percent of patients were, and 52 percent fear losing the human touch. Trust rises with transparency (a 65 percent comfort increase when physicians explain AI use), with experience, and most of all with physician endorsement; trust transfers from physician to tool the way a referral transfers from one doctor to another, and secrecy, not the AI itself, is the primary driver of distrust.

The forces ahead push in one direction. Agentic AI moves toward autonomous decisions, multimodal models fuse imaging, text, and structured data, and a projected shortfall of 10 million healthcare workers by 2030 makes adoption unstoppable. Gartner expects over 40 percent of agentic AI projects to be cancelled by the end of 2027 over cost and inadequate risk controls, but the shortage will drive AI in regardless of readiness, so the task is not to solve governance in the abstract but to build systems worth governing: techni-

cally excellent, validated, equitable, transparent, and honest about limits.

Watch out. Deploying AI to avoid hiring rather than to augment staff treats the symptom (not enough hands) while worsening the cause (not enough people willing to do the work).

Should You Build This?

This is the book’s last checkpoint, and the question is no longer whether the model works. It is whether the system should be deployed at all. **Build it when** you can name both the pathway and the fence: an FDA route with a PCCP covering the updates you plan to make, or a monitored structure like the Utah sandbox, where the pilot reports publicly and autonomy expands only if concordance holds above the agreed threshold. That is deployment as a supervised experiment, not a launch. **Think twice when** no one can answer the liability question, when the volume is too limited to establish safety (as Utah’s own first report conceded), or when governance is a slide rather than an infrastructure, since a board can only pause a

drifting model if the monitoring exists to detect the drift. **Who bears the risk:** formally the physician and the deploying organization; practically the patient,

who cannot tell whether the AI renewing their metformin was validated to the standard of the clinician it replaced.

Quick Self-Test

1. Name the three FDA pathways and the core limitation of 510(k) clearance.
2. What problem does a PCCP solve, and why do agentic systems strain it?
3. Give two ways the EU AI Act differs philosophically from FDA SaMD, and explain its extraterritorial reach.
4. What must a healthcare model card disclose that an aggregate accuracy number hides?
5. Why is drift monitoring the algorithmic analog of pharmacovigilance, and why does detected drift not trigger automatic redeployment?
6. Map the three-way liability distribution when an AI triage error harms a patient.

If You Remember One Thing

The distance between a working algorithm and a deployed clinical tool is not measured in lines of code. It is measured in regulatory clearance, governance structure, post-market monitoring, liability frameworks, and patient trust. Epic's sepsis model, Optum, and Watson failed not because the math was wrong but because no one had the authority, data, or mandate to stop them. Build the systems worth governing.

Richard J. Young, Ph.D. · richard@deepneuro.ai · Senior AI Research Scientist, UnitedHealth Group · UNLV Lee Business School · ORCID 0000-0002-1109-7552