The Coming Collision Between Foundation Models and Regulated Clinical Decision Support

JAN 21, 2026

Abstract

Foundation models are being deployed into clinical workflows at a pace that significantly outstrips the regulatory and safety frameworks designed to govern clinical decision support systems. This creates a structural mismatch between h we've historically validated, monitored, and assigned liability for clinical softwa versus how these probabilistic, continuously-evolving systems actually behave in production.

Key tensions explored:

- Why foundation models break classical CDS taxonomies (deterministic rules vs probabilistic reasoning)

- The validation paradox: static evaluation methods for dynamic, emergent syste

- Current regulatory arbitrage strategies and their likely expiration timeline

- Model drift as clinical drift: why statistical monitoring is insufficient

- Liability gaps when decision paths are non-human-readable

- Implications for product development, deployment strategy, and capital allocat

Table of Contents

Why Foundation Models Break Classica CDS Frameworks

The existing regulatory apparatus for clinical decision support was built around systems that behave in fundamentally predictable ways. A sepsis alert fires when specific lab values cross predetermined thresholds. A drug interaction checker queries a static database of known contraindications. A risk score applies fixed coefficients to a bounded set of input variables. These systems are deterministic transparent, and their failure modes are enumerable. You can trace every output to explicit rules or statistical weights that were frozen at the time of validation.

Foundation models operate under completely different principles. They don't ap rules or look up answers in tables. They perform approximate pattern matching billions of parameters trained on internet-scale corpora, then fine-tuned on dom specific data, then potentially further adapted through retrieval augmentation o prompt engineering at inference time. The same model given the same input on Tuesday might produce meaningfully different output on Thursday if the retriev context changed or if the model was updated overnight. There's no fixed feature The model doesn't "know" what clinical guidelines it's applying because those guidelines were absorbed probabilistically during training rather than encoded a explicit logic.

This creates immediate problems for how we classify and regulate these systems FDA's framework for software as a medical device relies heavily on whether the software "analyzes medical images or other medical data" to provide "time-critic information that influences clinical management. But foundation models blur th categories completely. Is a model that drafts differential diagnoses based on clin notes analyzing medical data or just reorganizing text? When it suggests additio workup based on subtle pattern recognition across thousands of similar cases it during training, is that clinical decision support or advanced autocomplete?

The traditional risk-based approach assumes you can bound the system's behavi defining what inputs it accepts and what outputs it produces. A chest x-ray class takes images in a specific format and outputs probabilities for a fixed set of pathologies. But a clinical foundation model might take discharge summaries, la trends, medication lists, nursing notes, and radiology impressions, then generate anything from prior authorization letters to treatment plans to patient educatio materials. The input space is essentially infinite and the output space is unconst natural language. You can't prespecify all the clinical scenarios it might encount all the ways its outputs might influence care.

The distinction between "decision support" and "information retrieval" collapse when the model is simultaneously doing both in ways that aren't separable. A se system that returns relevant journal articles when you query a diagnosis isn't ma clinical decisions. But what about a system that reads your entire clinical note, identifies the diagnostic uncertainty you haven't explicitly stated, retrieves evide about differential diagnosis strategies for that specific presentation, and then synthesizes that evidence into actionable recommendations? That's not pure retr anymore. It's also not purely algorithmic decision support because there's no algorithm in the traditional sense, just probabilistic text generation guided by le representations.

The deeper problem is that existing CDS frameworks assume clinical reasoning happens in human brains and software just provides inputs to that reasoning. Th clinician sees the alert, considers the context, and makes the decision. But found models are starting to do parts of the reasoning itself. They're identifying releva

clinical patterns you might have missed, weighing competing diagnostic hypothe integrating guideline recommendations with patient-specific factors. The line between "providing information" and "making recommendations" becomes meaningless when the system is doing natural language reasoning that looks indistinguishable from human clinical thought processes.

This matters enormously for how these systems should be validated and monitor Classical CDS validation involves testing on held-out datasets, measuring sensiti and specificity, and documenting performance across relevant subgroups. But w exactly are you validating when the system's behavior is emergent rather than programmed? You can test whether a foundation model correctly identifies seps historical cases, but that tells you nothing about whether it will give appropriate advice for a novel presentation it's never seen, or whether it will hallucinate contraindications that don't exist, or whether its recommendations will subtly d away from current guidelines as the underlying base model gets updated.

The Validation Paradox

Every medical device regulation assumes you can validate a system at a point in t clear it for market, and then monitor for problems during real-world use. The validation step is supposed to establish that the device works as intended across specified use cases. For foundation models in clinical settings, this assumption i structurally broken.

Traditional validation relies on data from the past predicting performance in the future. You train a pneumonia detection model on 100,000 chest x-rays, hold out 20,000 for testing, measure AUC and calibration curves, and declare the model r for deployment. This works because the model is frozen. The weights don't char The decision boundary remains constant. If real-world performance degrades, yo know something about the input distribution shifted and you can retrain.

Foundation models throw all of this out. First, they're often continuously update GPT-4 in January is not the same as GPT-4 in June. The base model might get refreshed monthly. Fine-tuning might happen weekly. Retrieval indices get upda

daily as new literature publishes. There's no single frozen artifact to validate. Ev you lock down the model weights, the system behavior changes because the retri context changes or because the prompting strategy gets optimized.

Second, foundation models exhibit emergent capabilities that weren't explicitly trained. A model trained primarily on general medical text might suddenly demonstrate reasoning about rare genetic syndromes it barely saw during traini because it learned deeper pattern recognition that generalizes. You can validate performance on your test set, but the test set fundamentally can't capture all the the model might be used or all the clinical scenarios it might encounter. The capability space is too large and too unpredictable.

Third, these models can fail in ways that don't show up in standard metrics. A m might maintain high accuracy on multiple choice questions while simultaneousl generating plausible-sounding but completely incorrect explanations. It might correctly identify the most likely diagnosis while missing critical safety contraindications. It might perform well on average but fail catastrophically on cases that are clinically important but statistically rare. Traditional validation lo aggregate performance. But clinical safety requires understanding worst-case behavior, and foundation models have long-tail failure modes that are nearly impossible to enumerate in advance.

The validation paradox is that the more capable these models become, the harde are to validate. A narrow model that only does one thing can be thoroughly teste that thing. But a general-purpose clinical reasoning model that can handle thous of different tasks needs to be validated across all those tasks, including novel combinations that weren't anticipated during development. You can't build a tes that covers "all possible clinical reasoning" any more than you can build a test se covers "all possible medical knowledge."

This creates a fundamental tension with regulatory expectations. Regulators wan prospectively defined endpoints, prespecified success criteria, and evidence that system performs as intended before it goes to market. But for foundation model intended" is deliberately open-ended. The whole point is flexibility and

generalization. Locking down specific performance characteristics defeats the purpose of using a foundation model instead of a narrow task-specific model.

Some teams are trying to thread this needle by validating specific clinical workfl rather than the underlying model. Instead of saying "this model understands medicine," they say "this implementation correctly triages chest pain patients" c "this deployment appropriately escalates sepsis alerts." They treat the foundatio model as a component in a larger system where the system-level behavior can be validated even if the model behavior is opaque. This is probably necessary but definitely insufficient. The system-level behavior is still dependent on model capabilities that can change unexpectedly, and validating one workflow doesn't t you anything about safety in adjacent use cases.

The monitoring challenge is even thornier. For traditional ML models, you watc distribution shift in inputs and degradation in standard metrics. For foundation models doing clinical reasoning through natural language, what do you even mo You can track user satisfaction scores but those might miss silent failures where model sounds confident but is wrong. You can have clinicians review a sample of outputs but sampling strategies assume you know what failure modes to look fo can compare model outputs to actual clinical decisions but that's confounded by clinicians potentially trusting the model too much or too little.

The deeper issue is that foundation models can develop new failure modes after deployment. A model might start hallucinating drug names that sound plausible don't exist. It might begin confidently asserting outdated guidelines because the were more prevalent in training data. It might pick up biases from the way clini interact with it, creating feedback loops where poor recommendations get reinfo None of these are detectable through traditional model monitoring because they not about statistical performance degradation, they're about the model learning in new ways.

Regulatory Arbitrage vs. Regulatory Inevitability

Right now there's a massive arbitrage opportunity in how foundation models are positioned to avoid medical device regulation. Every vendor building clinical AI acutely aware that crossing certain lines triggers FDA oversight, which means ye validation work and ongoing regulatory burden. So they carefully design produc stay on the "right" side of those lines.

The most common strategy is positioning the system as administrative rather tha clinical. Documentation assistants that turn voice notes into structured clinical Prior authorization tools that draft appeals letters. Scheduling systems that use natural language understanding to route patients. None of these are technically making clinical decisions, even though they're using the exact same foundation models and reasoning capabilities that could make clinical decisions. The distin is entirely about how the output gets framed and who the nominal decision-mak

Another approach is the clinician-in-the-loop defense. The system doesn't make decisions, it just provides suggestions that a qualified clinician reviews and appr This is how most clinical AI products are currently deployed. The model genera differential diagnosis but the doctor has to click through and confirm. The mode recommends medication adjustments but requires attending sign-off. The mode potential safety issues but leaves final judgment to the care team. In theory this the human responsible and the AI in a supporting role.

The problem is that this defense relies on assumptions about how clinicians actu interact with these systems that are increasingly questionable. Research on automation bias shows that humans tend to over-rely on algorithmic recommendations, especially when those recommendations are presented confid and the human is cognitively overloaded. A clinician reviewing fifty AI-generate summaries in an afternoon shift is not carefully evaluating each one against thei independent clinical judgment. They're pattern-matching for obvious errors and rubber-stamping the rest. The "clinician-in-the-loop" becomes a legal fiction rat than a meaningful safety control.

Some vendors are getting creative about task framing. Instead of "this model diagnoses pneumonia," they say "this model identifies imaging studies that requ

urgent radiologist review." Instead of "this model recommends antibiotics," they "this model surfaces relevant antibiotic stewardship guidelines for the current c context." The output is functionally similar but the framing emphasizes informa retrieval and workflow optimization rather than clinical decision-making. Whetl this distinction holds up under regulatory scrutiny is an open question.

The regulatory arbitrage is happening because the current framework wasn't de for systems that exist on a continuum from pure information tools to autonomou decision-makers. Foundation models can dial their level of clinical reasoning up down depending on how they're deployed. The same underlying technology can a simple documentation tool or a sophisticated diagnostic assistant. Vendors are rationally choosing deployments that capture value while minimizing regulatory exposure.

But this equilibrium is unlikely to be stable. Multiple forces are pushing toward eventual reclassification and tighter oversight. First, as these systems get more capable and more widely deployed, the gap between their nominal role and their actual influence on clinical decisions becomes impossible to ignore. When a foundation model is being used to draft treatment plans that get approved witho modification ninety-five percent of the time, calling it a "documentation assistar stops being credible.

Second, the inevitable patient safety incidents will force regulatory response. It's question of whether a foundation model will make a serious clinical error that h a patient, it's a question of when and how visible the case becomes. The first tim well-documented case emerges where a model hallucinated a contraindication th to a delayed diagnosis or recommended a treatment that caused preventable harı there will be enormous pressure to bring these systems under formal oversight.

Third, the regulatory agencies are actively working on frameworks for AI/ML-ba medical devices and software as a medical device. The FDA's predetermined cha control plans, algorithm change protocols, and good machine learning practice guidelines are all attempts to create pathways for continuously-learning systems

these frameworks mature, they'll likely expand to cover systems currently claimi be non-clinical. The direction of travel is clearly toward more oversight, not less

Fourth, liability pressure will push vendors toward clearer regulatory status. Rig now the legal landscape for AI-caused medical errors is unsettled. If vendors fac significant liability exposure for harms caused by their systems, they may actuall prefer to go through FDA clearance because it provides some legal safe harbor a clearer standards for what constitutes adequate safety validation. Being regulate burdensome but being unregulated and sued might be worse.

The smart money is betting that most current administrative and workflow applications will eventually be reclassified as clinical decision support or softwa medical device within three to five years. The question is whether vendors are building with that assumption or treating regulatory arbitrage as a permanent st Companies designing products that only work in the current regulatory gap are setting themselves up for expensive pivots when that gap closes. Companies buil with the expectation of eventual oversight, even if they're not there yet, are positioning for long-term defensibility.

When Model Drift Becomes Clinical Dri

Model drift in traditional ML systems is well-understood. Your fraud detection degrades because fraudsters adapt their tactics. Your demand forecasting model accuracy because consumer behavior shifts. You detect drift by monitoring inpu distributions and output metrics, then retrain on recent data to restore perform The underlying assumption is that the real-world phenomenon you're modeling changing while your model stays static.

Foundation models invert this. The model itself is changing through updates to models, fine-tuning on new data, refinements to prompting strategies, and expar to retrieval corpora. Meanwhile clinical medicine is also changing as guidelines update, new evidence emerges, and practice patterns evolve. You get drift on bot sides simultaneously, and they interact in ways that are extremely difficult to pre or monitor.

Consider a foundation model being used to help clinicians identify patients who might benefit from SGLT2 inhibitors for heart failure. The model was fine-tuned on clinical notes and guidelines from 2023. In early 2024, new trial data emerges showing benefit in a broader patient population. The clinical guidelines get updated within months. But the model's training data is frozen. It continues to suggest SGLT2 inhibitors based on the 2023 criteria while the standard of care has moved forward. This is clinical drift driven by medical knowledge advancing faster than model updates.

Now add model drift on top. The base foundation model gets updated to improve reasoning capabilities. The new version is better at multi-step clinical logic. But it was also trained on more recent internet text, which includes discussion of the updated guidelines. Now the model sometimes incorporates the new recommendations even though it wasn't explicitly fine-tuned on them. Except it does so inconsistently because it learned about the guidelines through unstructured web discussions rather than formal training. Its behavior has drifted in a direction that's partially aligned with current care standards but in unpredictable ways.

Traditional drift detection completely fails here. You could measure whether input note characteristics have changed, but clinical notes look statistically similar even as the underlying clinical context shifts. You could track model confidence scores, but foundation models are often poorly calibrated and confidence doesn't correlate well with correctness. You could monitor user feedback, but clinicians might not notice subtle drift in recommendations unless it causes obvious problems.

The deeper challenge is that drift in foundation models isn't just about statistical performance degradation. It's about the model developing new capabilities, losing old ones, and changing its reasoning patterns in ways that affect clinical appropriateness without necessarily affecting measurable accuracy. A model might become better at explaining its reasoning while simultaneously becoming more likely to hallucinate rare side effects. It might improve at handling complex cases while getting worse at recognizing when it should defer to specialist input. These changes don't show up in dropping AUC or rising calibration error.

Clinical language itself drifts in ways that create novel risks. New abbreviations common use. Terminology changes to reflect updated understanding of diseases. names get rebranded or biosimilars enter the market with similar names. A mod trained on 2023 clinical notes might not recognize terminology that became stan in 2025, or worse, might misinterpret new terms based on superficial similarity t ones. This is compounded by retrieval-augmented generation, where the model p information from external sources that are themselves constantly updating.

Feedback loops make everything worse. If clinicians start trusting the model's recommendations, their documentation might shift to align with what the model expects or suggests. The model then gets fine-tuned on this documentation, reinforcing its own patterns even if they're suboptimal. You get a drift spiral wh the model shapes clinical practice which shapes the training data which shapes t model. Breaking out of this requires external ground truth, but in many clinical scenarios there isn't a clear ground truth to validate against.

Some teams are trying to solve this through continuous evaluation on curated te that get periodically updated to reflect current guidelines and practice. This hel doesn't solve the fundamental problem. Test sets can't capture the full complexi clinical reasoning, and updating them requires clinical expertise that's expensive slow. By the time you've built a comprehensive test set for the current standard care, the standard of care might have shifted again.

Others are exploring adversarial testing where you deliberately try to find cases the model fails or produces outdated recommendations. This is valuable for find known failure modes but struggles with unknown unknowns. The model might developed new failure modes that your adversarial testing doesn't cover because didn't think to test for them. And adversarial testing at scale is expensive enoug most organizations only do it during major version updates, not continuously.

The most promising approaches involve some combination of continuous autom monitoring, periodic expert review of model outputs across diverse scenarios, systematic comparison against external clinical benchmarks, and explicit mecha for clinician feedback when model recommendations seem off. But even this is

insufficient because it's fundamentally reactive. You detect problems after they emerge rather than preventing them. For clinical systems where errors can cause patient harm, reactive monitoring isn't good enough but proactive prevention of seems nearly impossible with current technology.

The Liability Black Box

Medical malpractice law is built on a straightforward premise. A clinician has a of care to their patient. If they breach that duty through negligence and cause ha they're liable. The system works because clinical decision-making is human, and humans can be held accountable. Even when clinicians use tools and information systems, there's a clear chain of responsibility. The doctor ordered the wrong medication. The nurse missed the drug interaction warning. The radiologist mis the imaging study.

Foundation models break this chain because they introduce decision-making tha not fully attributable to any specific human. The model synthesizes recommenda based on learned patterns across billions of parameters. No individual can trace exactly why the model suggested what it suggested. The clinician who accepts th model's recommendation might not understand the reasoning well enough to cat errors. The engineers who built the model might not know what clinical knowle absorbed during training. The organization that deployed the model might not h visibility into how the base model was updated by its vendor.

When something goes wrong, who's liable? The most obvious answer is the clinic who acted on the model's recommendation. They're the licensed professional providing care. But this assumes they had meaningful ability to evaluate the recommendation independently, which becomes questionable as models get mor sophisticated. A foundation model that's correct ninety-nine percent of the time trains clinicians to trust it. When it fails on the one percent, expecting the clinic have caught the error is unrealistic.

Maybe liability rests with the vendor who provided the model. They're selling a product that influences clinical care. But foundation model vendors will argue t

providing a tool, not making medical decisions. They'll point to disclaimers stat the model should only be used by qualified clinicians exercising independent judgment. They'll note that they can't control how customers deploy or fine-tun base model. They'll claim the liability belongs to whoever implemented the spec clinical application.

Perhaps the health system that deployed the model is responsible. They decided integrate it into clinical workflows. They chose what validation to require before live. They trained staff on how to use it. But health systems will argue they relie vendor representations about safety and performance. They'll note they don't ha technical expertise to evaluate foundation models themselves. They'll point out t they implemented appropriate oversight processes like clinician review.

The problem is everyone has partial responsibility but no one has complete responsibility. The foundation model's behavior emerges from the base model tr (vendor), the fine-tuning data (health system), the retrieval context (could be eith the prompting strategy (clinical workflow designers), and the clinician's interact with the output. When the model hallucinates a contraindication, is that a traini data problem, a fine-tuning problem, a prompt engineering problem, or a clinici oversight problem? Probably some combination of all four.

Existing product liability frameworks struggle with this because they assume products have defined specifications and fail when they don't meet those specifications. A pacemaker that fails is a straightforward product defect. But foundation models are probabilistic. They're supposed to occasionally produce incorrect outputs. The question is whether a specific error represents acceptable model behavior or a defect, and there's no clear standard for making that determination.

Some legal scholars have proposed treating foundation model errors like physici errors and applying malpractice standards. The model would have a duty to mee standard of care for its specified clinical task. If its recommendations fall below standard and cause harm, the vendor is liable. This is conceptually clean but practically difficult because foundation models don't have professional training

licensure that defines a standard of care. What's the appropriate standard for an system that's more accurate than average doctors on some tasks and worse on otl

Another approach is treating foundation models like medical devices and applyi products liability law. If the model has a defect that causes injury, the vendor is strictly liable regardless of negligence. This gives patients a clear defendant and strong incentives for vendors to ensure safety. But it might also make foundatio model development prohibitively risky because the models will inevitably make and proving those errors weren't defects is nearly impossible.

A third option is creating a new liability framework specifically for AI in healthc potentially including no-fault compensation systems similar to vaccine injury programs. Patients harmed by foundation model errors get compensated withou having to prove fault. Funding comes from fees on vendors or health systems deploying the technology. This protects patients and removes some litigation ris it also removes some incentive for vendors to invest in safety since they're not fa unlimited liability exposure.

The messiest cases will be those where the model didn't exactly fail but influenc care in subtle ways that contributed to suboptimal outcomes. A model that prio likely diagnoses over dangerous ones, leading to delayed recognition of a rare bu serious condition. A model that confidently summarizes a patient history in a wa omits crucial details, causing subsequent providers to miss important context. A model that generates prior authorization justifications so effective that payers ap treatments the patient didn't actually need. None of these are clear-cut errors bu could contribute to patient harm.

Current reality is that most organizations are punting on these questions by req clinician sign-off on anything the foundation model suggests. This maintains the fiction that the clinician is the decision-maker and preserves existing liability allocation. But as models become more capable and their recommendations bec harder to independently verify, this fiction becomes more strained. At some poir we'll need actual answers to the liability questions rather than procedural workarounds that maintain ambiguity.

What This Means for Product Strategy and Capital Deployment

If you're building or investing in clinical AI, the coming collision between found models and regulatory frameworks should fundamentally shape your strategy. Companies that treat current regulatory ambiguity as permanent are setting themselves up for expensive pivots or outright failure when oversight inevitabili arrives. The winners will be those building for the world as it will be in five year as it is today.

First implication is that pure-play foundation model companies face regulatory that's hard to price. If you're building the next generation of clinical reasoning m as a platform play, betting that customers will handle regulatory compliance at t application layer, you're exposed to the risk that regulators decide platform prov need clearance regardless of how customers deploy the technology. This is espec true if your model is marketed for clinical use cases even if you technically don't control the deployment. The smarter approach is building regulatory strategy in platform from day one, even if current customers don't require it, because futur customers will.

Second implication is that application-layer companies need to decide whether t optimizing for current regulatory arbitrage or future regulatory inevitability. If clinical documentation tool is really doing clinical reasoning but framed as administrative support, you should be building the validation infrastructure you need when that framing stops working. This means prospectively collecting outc data, building automated safety monitoring, creating audit trails for model recommendations, and establishing clinical oversight processes. Doing this befo you're required to is expensive but avoids having to retrofit these capabilities un regulatory pressure when they're much more expensive and disruptive.

Third implication is that the value of narrow, validated clinical AI tools might increase relative to general-purpose foundation models. A foundation model tha do anything is hard to regulate because you can't validate "anything." A focused that does medication reconciliation or identifies imaging studies requiring urge

review has a defined scope that's much easier to validate against clinical standar
This suggests opportunity in building specialized foundation models that are sti
powerful but constrained to specific clinical domains where validation is tractab

Fourth implication is that companies building clinical AI infrastructure that hel
other teams deploy foundation models safely have a structural advantage. If you'
providing the monitoring, evaluation, validation, and compliance tooling that ma
foundation model deployment defensible, you're selling picks and shovels while
everyone else is panning for gold. This is especially valuable because most healt
systems and clinical software vendors don't have the ML ops expertise to build t
infrastructure themselves, but they're increasingly aware they need it.

Fifth implication is that integration and orchestration matter more than model
performance in isolation. A foundation model with slightly worse benchmark sc
but better audit trails, more predictable failure modes, clearer model cards
documenting training data and capabilities, and tighter integration with clinical
workflows will win over a more accurate model that's a black box. Health system
buying clinical AI are increasingly sophisticated about deployment risk versus m
risk. They know that the best model on paper often isn't the best model in produ

Sixth implication is that companies building clinical AI need clinical expertise o
founding team or in senior leadership, not just as advisors. The regulatory and s
challenges aren't purely technical problems that engineers can solve. They requi
deep understanding of clinical workflows, medical liability, healthcare complian
and how clinicians actually make decisions under uncertainty. Teams that treat
clinical input as a nice-to-have rather than core to product development are buil
products that won't survive contact with real clinical environments.

Seventh implication is that incumbents with existing regulatory relationships an
compliance infrastructure have meaningful advantages over startups in highly
regulated clinical use cases. A startup building a foundation model for clinical
decision support is facing a multi-year regulatory pathway with significant capit
requirements. An incumbent EHR vendor or clinical software provider building
same capability can leverage existing FDA relationships, quality management sy

and clinical validation expertise. This suggests that for high-risk clinical applica the winning strategy might be building technology that gets acquired by incumb rather than trying to go to market independently.

Eighth implication is that international markets with less developed AI regulati might seem like attractive near-term opportunities but create long-term risk. If build a clinical AI product optimized for markets with minimal oversight, you're building organizational muscle memory around moving fast and breaking things healthcare, which is exactly wrong for eventual developed market entry. Better t build for the hardest regulatory environment first and expand from there than t build for easy markets and try to add rigor later.

Final implication is that the companies that win will be those that view regulato compliance as a feature rather than a cost center. When every vendor is using sir foundation models and achieving similar performance, the differentiation comes safety, reliability, auditability, and trust. Building superior validation infrastruc clearer model documentation, better drift monitoring, and more robust clinical oversight becomes your moat. This requires upfront investment that many startu will skip because it doesn't show up in demos or benchmark comparisons. But it what enterprise healthcare buyers increasingly care about as they get burned by vendors who oversold capabilities and underinvested in safety.

The next few years will separate clinical AI companies into two groups. One gro will keep pushing boundaries of what foundation models can do while treating s and regulation as problems to solve later or route around. These companies will fast, show impressive capabilities, and attract lots of attention. Some will succee spectacularly if they time regulatory changes correctly. Most will hit regulatory or safety incidents they're not prepared for.

The other group will build foundation model applications with regulation and s as first-order constraints rather than afterthoughts. They'll move slower in the n term. Their products might seem less impressive in demos because they're hedg conservative. But they'll build defensible businesses that survive the transition f

regulatory ambiguity to regulatory clarity. They'll be positioned to scale when he systems and payers get serious about clinical AI governance.

The capital opportunity is identifying companies in the second group before the market fully prices in the regulatory inevitability. Right now there's still alpha in backing teams that are building for the regulated future rather than the unregul present, because most investors are still chasing impressive demos over sustaina compliance infrastructure. That window is closing as the regulatory framework solidifies and safety incidents start accumulating. But for now there's still meani mispricing between what clinical AI companies should be worth based on their current capabilities versus what they'll be worth based on their ability to naviga coming regulatory collision.

3 Likes

← Previous

Nex

Discussion about this post

Comments Restacks

Write a comment...