

The OpenAI Anthropic AI Arms Race Pivoted From Models To Services & Deployment. And Healthcare Is The Stress Test.

How OpenAI's PE-Backed Deployment Co And Anthropic's Blackstone-Hellman-Goldman JV Are Fighting To Own The Services Layer

MAY 06, 2026



Share

Quick Links: Knowledge Base, Podcast, and Social

Knowledge Base — search and filter every article and podcast episode by topic, section, and keyword: kb.onhealthcare.tech

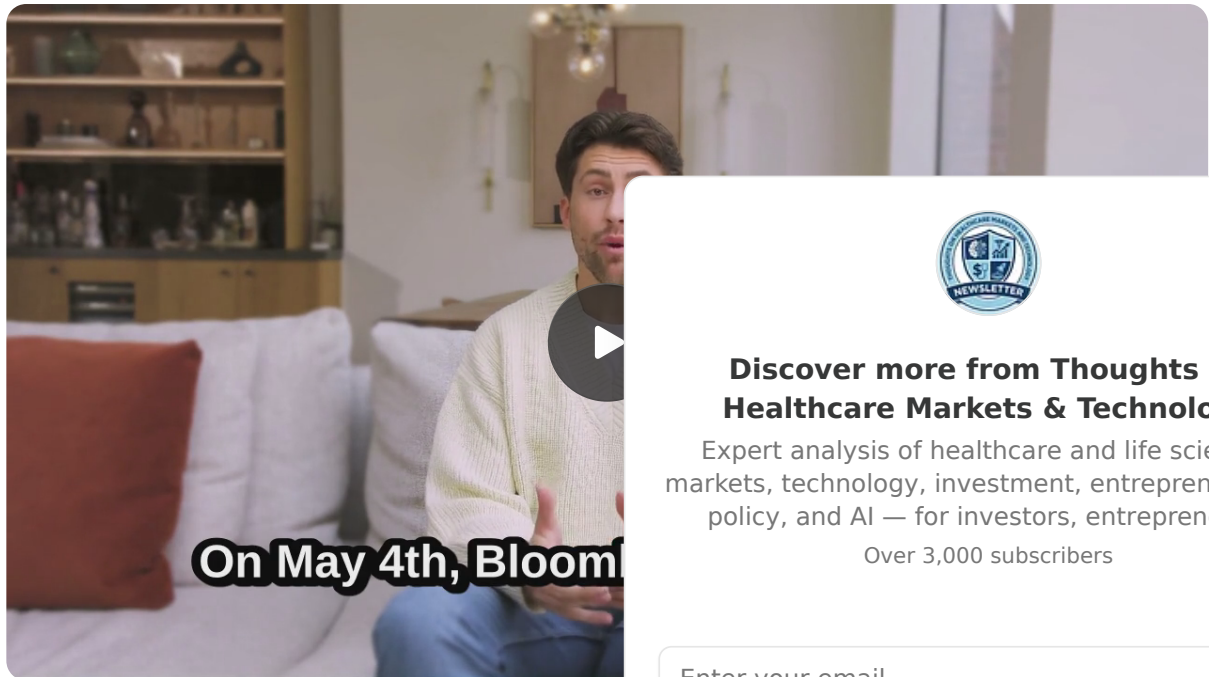
Listen to the Podcast — every article is also available as an audio episode. Free subscribers get the public episodes; paid subscribers get the full archive including subscriber-only episodes. Listen on [Apple Podcasts](#), [Spotify](#), or browse all episodes on the [Substack Podcast page](#).

For paid subscribers — your subscription unlocks the entire research archive (538+ deep-dives), every paid podcast episode, and full search inside the Knowledge Base. To listen to paid episodes in Apple or Spotify, link your Substack subscription via the show settings on those platforms (instructions inside the Substack app under Subscriptions → Podcast).

For free subscribers — free posts and free podcast episodes are always public on Apple/Spotify and Substack. Upgrade any time at onhealthcare.tech/subscribe to access the paid archive and paid episodes.

Follow on Social — [X](#) · [YouTube](#) · [TikTok](#) · [Instagram](#)

Video Preview



Part I Podcast free on [Podcasts](#), and the Web P



Thoughts on Healthcare Markets & Technology



Part I: The OpenAI Anthropic Arms Race Pivoted From Models To Services & Deployment. And Why Healthcare Is The Stress Test.

Two AI labs. Two consecutive days. Two PE-backed deployment vehicles. Same thesis. Here is what is actually happening and why healthcare is the stress test nobody was ready for...

 Listen now

6 days ago · Thoughts on Healthcare



Discover more from Thoughts on Healthcare Markets & Technology

Expert analysis of healthcare and life sciences markets, technology, investment, entrepreneurship policy, and AI — for investors, entrepreneurs,...

Over 3,000 subscribers

Enter your email...

Subscribe

By subscribing, you agree Substack's [Terms of Use](#), and acknowledge its [Information Collection Notice](#) and [Privacy Policy](#).

Already have an account? [Sign in](#)

Part II Podcast episode for paid subscribers only. Also available on Spotify.



Thoughts on Healthcare Markets & Technology



Part II: The OpenAI Anthropic Arms Race Pivoted From Models To Services & Deployment. And Why

Healthcare Is The Stress Test.

Two AI labs. Two consecutive days. Two PE-backed deployment vehicles. Same thesis. Here is what is actually happening and why healthcare is the stress test nobody was ready for...

▶ Listen now

6 days ago · Thoughts on Healthcare

Abstract

1. Bloomberg reported May 4 that OpenAI is finalizing a roughly \$10B JV with PE firms to deploy AI inside enterprises. Anthropic announced a roughly \$1.5B JV with Blackstone, Hellman & Friedman, and Goldman Sachs the next day. Two days, two parallel structures, same thesis.
2. Both labs have arrived at the same conclusion. Model capability is no longer the binding constraint on enterprise value capture. Deployment is. The labs are spinning up the implementation layer rather than trying to grow it inside their core orgs.
3. Healthcare is the worst-case stress test for that thesis. EHR plus claims plus RCM integration, HIPAA plus HTI-1 plus FDA exposure, clinical validation, change management, and data fragmentation across payer, provider, and life sciences ecosystems all sit on top of the model call. Generic enterprise rollout playbooks die here.
4. PE is the distribution channel, not a passive backer. Blackstone, Bain, Advent, KKR, Welsh Carson, GTCR, and Hellman & Friedman collectively control physician rollups, RCM platforms, CROs, pharma services, home health, behavioral health, and payer services vendors. That portfolio is a built-in deployment substrate that bypasses the slow path through health systems.
5. The regulatory floor is non-trivial. ONC HTI-1's DSI requirements, NIST AI RMF GenAI profile, HHS AI Strategy, FDA predetermined change control plan guidance, and a growing patchwork of state laws wrap every clinical or coverage-decision model call in a multi-stack governance environment.
6. Likely winners are not the labs themselves and not pure SaaS wrappers. They are orgs that own integration, evals, orchestration, audit and provenance, and forward-deployed engineering plus clinical informatics. Services becomes the control layer.
7. The read for builders, operators, and capital allocators: stop benchmarking models. Start mapping where deployment friction creates margin.

Table of Contents

1. Video Preview

2. Podcast, Part I (Free)

Thoughts on Healthcare Markets & Technology is a reader-supported publication. To receive new posts and support my work, consider becoming a free or paid subscriber.

3. Why the model wars are functionally over

4. What OpenAI and Anthropic actually just did

5. Deployment is the real bottleneck

6. Healthcare as the hardest battlefield

7. The PE distribution play, hiding in plain sight

8. The regulatory floor nobody at the model layer is engineered for

9. Services as the control layer

10. What builders, operators, and capital allocators should do now

11. The closing read

Why the model wars are functionally over

Punchline first. The frontier model gap is collapsing. Not in some absolute philosophical sense, the labs are still doing real research, but in the pragmatic sense that matters to anyone trying to ship into a regulated workflow. GPT 5.x, Claude Opus 4.7, Gemini 3, Grok 5, even DeepSeek's open weights stuff, they all clear the bar for the vast majority of enterprise tasks: summarization, structured extraction, code gen, function calling, reasoning over long context. The differences at the top are real but they are the kind of differences that show up in benchmarks and in narrow agentic eval suites, not in whether a prior auth workflow actually closes.

And the people running those workflows have figured this out. Walk into any decent CIO office at a national health plan, a top 25 IDN, a mid-cap biotech, or a PE-backed RCM platform, and the conversation is not “which model is best.” It is “we have access to all of them, the 13 month POC produced exactly one production deployment, and that deployment is currently being audited by compliance because we cannot trace the provenance of three of its outputs.” The bottleneck is not on the inference side. It never really was, except briefly in late 2022 when nobody had figured out function calling and everybody thought RAG was the answer to everything.

So when both OpenAI and Anthropic, on consecutive days in early May, announced massive PE-backed services and deployment vehicles, this is not a coincidence. It is two of the most strategically clear-eyed companies in the space arriving at the same conclusion at roughly the same moment: the labs do not capture enterprise value by selling tokens. They capture it by owning the implementation layer where tokens get turned into revenue.

That conclusion is uncomfortable for a lot of the AI economy because it implies the value isn't actually in the model. It is in the consulting, the integrations, the evals, the change management, the audit logs, the compliance frameworks, the clinical informatics work, all the unsexy stuff that traditional enterprise software companies have been doing for thirty years. Which means the model providers, if they want to capture more than commodity margin, have to become services companies. And that, as anyone who has ever tried to build a services business inside a product company can tell you, is a really hard internal pivot. So they did the next best thing. They externalized it through PE.

What OpenAI and Anthropic actually just did

The structures are different but the strategic intent rhymes hard. Per Bloomberg's May 4 report, OpenAI is finalizing what looks like a roughly \$10B joint venture with private equity backers to deploy AI inside enterprise environments. The framing is not “consulting.” It is closer to a forward-deployed engineering arm with capital, headcount, and a mandate to own implementation across the buyers OpenAI cannot service directly. The Reuters report that landed the next day suggested both OpenAI and Anthropic have been in talks to acquire AI services firms, which fits, you cannot stand up a forty thousand person deployment org organically when your best engineers want to do model research.

Anthropic announced its own version one day later. Per Anthropic's release, plus WSJ's reporting on the capital structure, the JV brings Blackstone, Hellman &

Friedman, and Goldman Sachs together at roughly \$1.5B in committed capital. The pitch is similar: a dedicated entity that acts as the deployment muscle, taking Claude into enterprise environments, with a particular emphasis on the kinds of regulated, complex, integration-heavy workflows where model access alone does not produce ROI. Blackstone in the press release was explicit about portfolio deployment, this is not a passive financial play, it is a distribution engine.

The choice of partners on the Anthropic side is the most interesting tell. Blackstone is one of the largest PE platforms in healthcare, with significant exposure to physician practice rollups, IT services, life sciences contract research, and provider services. Hellman & Friedman has a long bench of healthcare and financial services portfolio cos. Goldman Sachs brings the financial services distribution and the M&A muscle. Pull the lens back, and what you have is a coalition that can plausibly say, with a straight face, that they have direct or one hop access to a meaningful slice of the Fortune 500 plus a dense cluster of mid-market healthcare assets. That is exactly the right shape if your strategic problem is, “we have a model people want to use, and an enterprise sales motion that takes 14 months.”

Here is the part most coverage missed. Both vehicles are not quite consulting firms in the McKinsey sense. They are closer to what you would get if you fused a forward-deployed engineering org (think Palantir’s FDE model, scaled) with a PE operating partner team. The McKinseys, BCGs, Accentures, and Deloitte of the world have already been generating real revenue selling GenAI advisory and pilot work, that horse left the barn in 2024. What these new vehicles offer is something the consultancies cannot easily match: a privileged relationship with the lab itself, including access to model roadmaps, custom fine-tunes, dedicated capacity, and integration patterns the rest of the market does not see for months. That is a real moat if you can hold it.

Whether they can hold it is a separate question. The track record of model providers trying to run services arms is mixed. IBM’s history with Watson Health is the obvious cautionary tale. The structural fix here, externalizing the services work into a separately capitalized PE-backed entity, is intended to address exactly the cultural and incentive mismatch that broke Watson Health, but a structure does not by itself solve the problem of doing implementation in a regulated environment where the failure modes are not “we missed quarter.”

Deployment is the real bottleneck

If you are a healthcare buyer who has been around for any enterprise software cycle in the last twenty years, this is not news. It was true for EHR rollouts, for population

health platforms, for the entire prior auth automation cohort, for clinical decision support, for ambient scribes, for revenue cycle automation. The story is always the same. The technology is, by year three, basically fine. The deployment is where the company lives or dies.

What is genuinely new with frontier models is that the technology has finally outpaced the rate at which buyers can deploy it. That is the inversion. For most of enterprise software history, the constraint was “the tool cannot do what we need.” Now the constraint is “the tool can do far more than our org can absorb.” Look at any decent ambient scribe rollout. The model can transcribe and structure a clinical encounter at near-physician quality. The deployment friction is the EHR write-back, the template alignment with each specialty, the billing code mapping, the medico-legal documentation policy, the physician training, the QA loop, the audit trail, the malpractice insurance carrier sign-off. Each one of those is a project. None of them is solved by a better model.

Multiply that across every workflow where AI is actually getting traction. Prior auth automation has the same shape, the model is plenty smart, the deployment is owning the X12 EDI integration plus the payer-specific medical policy parsing plus the appeals workflow plus the audit trail when the algorithm denies. RCM denials management, same shape, smart enough model, deployment is the 837 and 835 plumbing plus the work queue ergonomics plus the human-in-the-loop UI plus the clearinghouse contracts. Clinical trial site selection, smart enough model, deployment is the protocol parsing plus the IRB workflow plus integration with CTMS plus the regulatory submission overlay.

What OpenAI and Anthropic just announced is, in plain English, an attempt to industrialize the answer to that deployment problem at scale. Not “we will sell you a model and good luck.” Closer to “we will bring a team, capital, integrations, evals, and a multi-year operating relationship, and we will turn the model into a productive worker inside your environment.” That is an entirely different business than selling tokens. It has different unit economics, different sales motion, different talent profile, and very different ROI proof requirements.

In healthcare specifically, that pivot maps almost perfectly onto how buyers already think. Health systems do not buy software. They buy outcomes, services, and operating models, often wrapped in a tech stack but rarely independent of one. Payers are the same. Pharma is closer to the traditional software buyer but increasingly thinks in terms of full-stack research operations. The deployment-first frame is, awkwardly for some of the AI-native pure plays, just how healthcare has always bought.

Healthcare as the hardest battlefield

Here is where the analysis gets concrete for anyone reading this in a healthcare seat. Every dimension that makes enterprise AI hard is dialed to maximum in healthcare, simultaneously.

Workflow integration. The average mid-sized health system runs Epic or Oracle Cerner, plus a portfolio of bolt-on systems for radiology, lab, pathology, billing, scheduling, patient comms, population health, and roughly fifteen ancillary point solutions, half of which were acquired during a prior platform consolidation and half of which are running on dialects of HL7 v2 with one engineer in the basement who knows the interface. Plug in claims data and you add 837, 835, 270, 271, and 278 transactions plus FHIR R4 endpoints with varying degrees of conformance. Plug in pharma and you add CTMS, EDC, ePRO, and a regulatory submissions stack. The model is the easy part. The integration is two engineering years per workflow per environment, and the workflow is rarely the same across two health systems.

Regulatory exposure. HIPAA at the floor, but HIPAA is the easy regulatory layer because the playbook is mature. The harder layer is ONC's HTI-1 final rule, which now requires algorithm transparency disclosures (the DSI requirements, decision support intervention) for predictive AI inside certified health IT. That alone wraps a real governance stack around any model that influences a clinical decision. Then layer in FDA's evolving framework for software as a medical device, particularly the predetermined change control plan guidance for adaptive ML, which materially changes how you can deploy a model that updates over time. Then add state-level rules, California's law on AI in utilization management decisions being one of the more aggressive examples. None of this gets simpler.

Validation and clinical safety. Saying a model is "safe" in a generic enterprise context means it does not expose PII, does not hallucinate confidently in customer-facing chat, and has reasonable evals on a relevant benchmark. Saying a model is "safe" in clinical context means it has been validated against ground truth on the population it will encounter, with attention to demographic subgroup performance, calibration of its uncertainty estimates, drift monitoring, and a meaningful escalation path when it falls outside its operating envelope. HealthBench and similar evaluation frameworks are a real attempt to make this measurable, but evals are an input to validation, not a substitute. Validation is local, ongoing, and expensive.

Change management. Clinicians are not anti-tech, they are tired and skeptical, often correctly. Burn one med staff meeting with a half-baked AI tool that triggers more after-hours work and the rollout is dead at that site for two years. The single biggest

predictor of whether an AI deployment in a clinical setting succeeds is whether the org has competent clinical informatics leadership willing to do the unglamorous work of redesigning the workflow around the tool, rather than dropping the tool on top of a workflow that was already broken.

Data fragmentation. The data needed to make any of this work is split across payer claims, provider EHRs, pharmacy data, lab systems, device telemetry, social determinants, and patient-reported outcomes, sitting in roughly twenty different formats and two dozen consent regimes. The labs, even with the world's best models, are not going to fix that data problem from the API side. Somebody has to do the work of stitching it together, governing it, and making it usable inside a model-augmented workflow. That somebody, increasingly, looks like the deployment company.

This is the core reason healthcare is the stress test. A generic enterprise deployment can fail and cost a customer a few million dollars and a quarter. A healthcare deployment can fail and cost a patient. The bar is not the same. The labs know this, which is partly why both of them are layering specialized vehicles, evals, and partner ecosystems around their core API. They cannot, from inside a foundation model lab, do the safety case work for every clinical workflow at every health system. So they need a deployment layer that can.

The PE distribution play, hiding in plain sight

The most under-analyzed angle in the coverage of these announcements is the PE side. The press treated it as financial backing, which is half right and entirely insufficient. Private equity in healthcare is not a passive checkbook. It is the largest, most aggressive distribution channel for healthcare technology that exists outside of the federal government.

Consider what PE actually owns in healthcare today. On the provider side, Blackstone, KKR, Bain, Advent, Welsh Carson, GTCR, and Hellman & Friedman collectively control or have major positions in physician practice rollups across primary care, dermatology, ophthalmology, cardiology, gastro, anesthesia, ortho, urology, oncology, women's health, and behavioral health. Hundreds of practice locations per platform, often with fifty to two hundred million dollars of revenue per platform, sometimes more.

On the services side: RCM platforms, healthcare IT consultancies, claims editing vendors, prior auth services bureaus, coding companies, denials management firms, healthcare analytics platforms. Pharma services is its own animal: CROs (where

companies like Parexel, ICON, and others have been bouncing between PE and public ownership), site networks, patient recruitment cos, real world evidence vendors, regulatory affairs consultancies, manufacturing services. Home health has been a major PE vertical for the last decade, despite the Medicare Advantage reimbursement turmoil. Behavioral health rollups, particularly post-pandemic, are a significant footprint. Payer services vendors, particularly the ones serving plans on Medicaid managed care contracts, are mostly owned by PE.

The strategic implication that few people have made explicit: PE-backed healthcare service platforms are, in aggregate, the single largest body of healthcare workflow surface area where AI deployment can be standardized at speed. Health systems, by contrast, are slow, internally heterogeneous, and operate on multi-year capital cycles that typically lag tech adoption by two to four years. A national radiology PE rollup with a hundred sites running on a normalized PACS and a centralized billing operation is, from a deployment POV, a much easier substrate to deploy AI into than a hundred independent radiology groups would be. PE has been doing the rollup work for fifteen years, and AI is, conveniently, arriving just as those platforms have hit a maturity point where they need a productivity step function.

Now read the JV announcements again with that frame. Blackstone's involvement on Anthropic's side is not capital. It is access to a portfolio of healthcare and adjacent service businesses where Claude can be deployed in an environment where the platform-level economics, governance, change management, and integration work has either already been done or can be done much faster than at a comparable health system. Same logic on Hellman & Friedman and Goldman, with different industry mixes. OpenAI's PE-backed deployment co, judging from the reporting, is built around the same insight, just less explicitly tied to specific PE houses' portfolios so far.

This produces a closed loop that is genuinely novel in healthcare's adoption history. The lab builds capability. The PE firm uses its operating partner muscle to push deployment across portfolio cos. The deployment generates real workflow data, used in turn to refine the model and the implementation playbook. Economic value compounds at the portfolio level, since EBITDA expansion in PE-backed cos is the explicit thesis. And the next portfolio company, or the next acquisition, inherits a more battle-tested playbook than the previous one.

The pace implication is significant. If this thesis plays out, the first wave of "AI native" healthcare margin expansion will not show up at academic medical centers or large publicly traded health systems. It will show up in PE-backed service platforms,

where the playbook can be standardized across forty similar sites and the operating partner can force adoption in a way that no health system COO ever could.

The regulatory floor nobody at the model layer is engineered for

So the technology is fine and the distribution exists. What about the legal and regulatory layer that sits underneath? This is where the model providers' bare API really cannot serve. It has to be wrapped.

Start with HTI-1, the ONC final rule that became enforceable in early 2025. The DSI requirements impose specific obligations on certified health IT developers to disclose attributes of predictive algorithms used in clinical decision support, including evidence basis, intended use population, performance metrics, and warnings about limitations. Pulled in one direction, this is a healthy transparency floor. Pulled in another, it is a non-trivial documentation burden that effectively becomes a barrier to entry for anyone deploying a model behind a clinical workflow. The rule frames the obligation around the certified health IT developer, but the practical compliance work has to happen at the deployment layer, where the model is actually being used.

NIST's AI Risk Management Framework, plus the GenAI profile, has become a de facto baseline for enterprise governance, including in healthcare. It is not a regulation, it is a framework, but increasingly customers and procurement teams are asking vendors to map their controls to it. Doing that mapping, and producing the artifacts, is again deployment work. It is not something the API does for you.

HHS released its overarching AI strategy laying out federal posture across CMS, FDA, ONC, NIH, HRSA, and the broader department. The strategic thrust is roughly: encourage adoption while standardizing governance. CMS specifically has signaled increasing scrutiny of AI used in coverage and utilization management decisions, particularly in Medicare Advantage post the well-publicized denials rate concerns. That scrutiny is going to manifest as documentation requirements, audit trails, and probably eventually new conditions of participation language. Whoever is deploying your prior auth automation has to be able to produce that documentation on demand. Most foundation model APIs cannot.

FDA's evolving framework for AI-enabled device software is a separate but adjacent stack, particularly the predetermined change control plan guidance, which lets manufacturers pre-specify how a model can update over time without each update triggering a new submission. This is an attempt to reconcile the regulatory need for stability with the technical reality that models drift and improve. It is also an

enormously detailed engineering and governance stack to actually implement correctly.

State law adds another layer. California, Texas, Colorado, Illinois, and New York each have variations of laws or proposed legislation governing AI use in healthcare, particularly in utilization management and employment contexts. The current patchwork is the kind of thing that consumes enormous compliance effort and that no foundation model API is going to abstract for you. Multistate health plans, multistate provider groups, and any vendor selling into them have to navigate this themselves.

Pulled together, the regulatory floor in healthcare is not simply “there are some rules.” It is a multi-stack governance environment where the model is one of perhaps fifteen artifacts that have to be produced, documented, and maintained for any production deployment. The deployment company bet, in healthcare specifically, is partly a bet that owning this governance layer is more defensible than owning the model. If that turns out to be true, and there is no obvious reason it will not, the value capture geography changes meaningfully.

Services as the control layer

Here is the part that reframes the bull case for healthcare AI investing.

The default mental model has been “model providers at the top, vertical SaaS startups in the middle, point solutions at the bottom.” Most healthcare AI investing in 2023 and 2024 was driven by that picture. The deployment company thesis breaks the picture. If services are where deployment happens, then the company that owns services owns the customer, owns the data, owns the workflow, owns the audit trail, and ultimately owns the economic surplus.

The new control layer looks like this. At the bottom is the model, increasingly commoditizing on price and capability across at least four credible providers. Above that is the integration plane, EHR, claims, RCM, pharmacy, devices, CRM, ITSM, etc., which is highly proprietary, slow to build, and where the actual data leverage lives. Above that is the evaluation and validation plane, the equivalent of HealthBench but extended into custom evals for each specific workflow with each specific customer’s data, the local validation that matters for clinical and regulatory sign-off. Above that is the orchestration plane, where models, retrieval, tools, and human-in-the-loop steps get composed into workflows. Above that is the audit and provenance plane. And at the top is the forward-deployed engineering and clinical informatics function that does the bespoke implementation work.

A pure model provider lives on one of those layers. A traditional SaaS vendor occupies maybe two. A deployment company occupies all of them, by design. That is the strategic point. Whoever owns the most layers in this stack accrues the most value, because the layers above the model are stickier (multi-year integration debt), more defensible (data and process moats), and increasingly more regulated (governance and audit requirements).

A useful precedent is the way Palantir moved up the stack from generic data platform to forward-deployed implementation in defense, then in healthcare via its Foundry deployments at HCA and others. The economics of that model are well known by now: high gross margins on the platform, plus services attach that grows the lifetime value far beyond what a pure software ARR model can. Apply that pattern to a model-native business, with PE-backed distribution, and the value capture potential is significant.

For builders, the practical implication is that pure-play model layer companies that try to go it alone in healthcare are going to find themselves displaced or absorbed by deployment-led entrants that come with capital, integration partners, governance frameworks, and the implicit endorsement of one of the labs. Pure SaaS startups will face the same pressure. The startups that survive in this environment are the ones whose product is, in essence, a deployment company in a trench coat: heavy on services attach, deep customer relationships, tightly integrated to a specific workflow, and producing data and process artifacts that can be defended.

What builders, operators, and capital allocators should do now

Three sets of recipients here, three different reads.

For builders, especially seed and Series A founders building healthcare AI, the strategic move is not to compete with the lab plus PE plus consulting army. That is a war you cannot win on capital. The move is to occupy a workflow surface area that is too narrow, too clinically specific, or too operationally weird for a general-purpose deployment co to bother with, while building the data assets and integration depth that make you a credible acquisition target later. Specialty-specific workflow tooling, surgical and procedural data, niche revenue cycle problems with significant margin pools, regulated decision points where the customer needs a vendor with skin in the game on outcomes, all of these are durable. Generic copilots, shallow EHR overlays, and undifferentiated clinical chatbots are cooked.

For operators in PE-backed healthcare service platforms, this is one of the cleanest opportunities to compound EBITDA in a long time, and it is going to favor

management teams that move fast. Build the internal AI program now, before the deployment cos arrive with their roadshow decks. Put a senior person on it, ideally a former clinical or operations leader with credibility, not a VP who is bored. Identify three to five workflows where the operational data is good enough to drive a real proof of value. Negotiate hard on data rights when partnering with deployment vendors, because the data leverage is what you actually have and you should not give it away. And remember that the relevant comparison is not “us versus the AI.” It is “us versus the next portfolio company that does this better.”

For health system C-suites, the read is harder. The deployment cos will arrive selling a centralized, opinionated stack. Some of that pitch is real, some of it is overconfident. The genuinely useful posture is to set up a small, technically credible internal evaluation team that can pressure test what the deployment cos bring, not a big committee. Invest in clinical informatics, since whatever you deploy will fail without it. And resist the temptation to standardize prematurely on a single foundation model partner, because the negotiating leverage from optionality matters more this year than next.

For capital allocators, the questions to ask the next time a healthcare AI deal lands on the desk are different than they were in 2023. Stop asking which model the company uses. Ask what integrations they own, what data exclusivity they have, what their forward deployed engineer to seller ratio is, what the governance and audit posture is, how their evals look on real customer data, and whether the company has a coherent answer for a world in which the foundation model is cheaper next year, twenty percent better, and not a differentiator. The answers to those questions are how to tell a defensible deployment company from a model-wrapped SaaS company in a worse cap structure.

A more general read for everyone. Watch the consultancies, Accenture, Deloitte, IBM, Cognizant, Infosys, Wipro, because they are the obvious targets for the kind of acquisitive tucking-in that the Reuters report alluded to, and any serious capability move by either lab plus PE coalition will likely involve absorbing or partnering with a portion of that workforce. Watch Epic, because Epic owns a meaningful piece of the integration layer the deployment cos need, and Epic’s posture toward third party AI is not a passive consideration. Watch the FDA, because the predetermined change control plan workstream is going to mature over the next twenty four months and that will materially affect what kinds of products can be productized into devices versus kept as decision support.

The closing read

The model wars made for great content. Benchmark drops, leaderboards, scaling debates, capability evals, the whole thing. It was fun. It is not where the next phase of value capture happens.

The next phase is messier and slower. It is twelve month integration projects, contract redlines on indemnification language, hour long change advisory board meetings, and the very specific kind of organizational pain that comes from getting a clinician to actually trust the output of a model when their license is on the line. None of that fits in a tweet. All of it is where the dollars actually flow.

OpenAI and Anthropic, by spinning up PE-backed deployment vehicles two days apart, just made the structural bet visible. They are saying, in a way that anyone who reads cap tables can see clearly: the model is the loss leader. The deployment is the business. Whether they execute on that bet is unsettled, neither of these vehicles has booked meaningful enterprise revenue yet, and the implementation muscle they need to build from a standing start is genuinely hard to scale. But the read on the strategic geography is correct.

For healthcare specifically, this is going to be the most consequential structural shift in healthcare AI since the original ChatGPT moment. Not because the models are getting better, although they are. Because the distribution and deployment stack is finally getting capitalized at the scale that actually matches the size of the workflow surface area. The next two years will not be decided by which lab has the smartest model. They will be decided by which deployment coalition can move fastest through a few thousand PE-backed service platforms, a few hundred plan-side procurement processes, and the governance scaffolding that wraps both.

Tldr for anyone reading this from a healthcare seat: stop benchmarking models. Start mapping where deployment friction creates margin. The next decade of value capture in this industry is going to be built on top of that map.

Thoughts on Healthcare Markets & Technology is a reader-supported publication. To receive new posts and support my work, consider becoming a free or paid subscriber.



3 Likes

[← Previous](#)

[Next →](#)

Discussion about this post

Comments

Restacks



Write a comment...

