The AI clinical infrastructure compan
why the real money in Health AI isn't
the models

FEB 22, 2026 • PAID

5 2 Share

Abstract

This essay makes the case that the most durable and defensible business in healt over the next decade isn't building foundation models – it's building the deploy governance, and validation infrastructure that makes those models safe to use in clinical settings. The argument draws on the collective action problem facing he systems, the lessons of enterprise infrastructure companies like Azure, and the structural advantages of coalition-based data networks in regulated industries.

Key claims:

- Foundation model competition is already a race to the bottom among the worl best-capitalized companies. Health tech entrepreneurs don't need to be in that r

- Health systems lack the internal capability to deploy AI safely and at scale. Thi a gap a few vendors are going to close – it's a systemic infrastructure deficit.

- The clinical deployment and governance layer is the Azure of health AI. It sits between the model and the bedside and does the unglamorous, high-margin wor making AI actually work.

- Coalition-based deployment networks create compounding data and validation advantages that point solutions can't replicate.

- This is a venture-backable company with a realistic path to $500M+ ARR.

Table of Contents

The Setup: Why Health Systems Can't This Themselves

Start with a number: there are roughly 6,000 hospitals in the United States, oper across about 900 health systems of meaningful size. Every single one of them is g to need to figure out AI over the next decade. Not because they want to, necessa but because the economics of healthcare delivery are going to force their hand. I is the biggest cost driver in hospital operations – somewhere north of 55-60% of expenses – and AI is the only plausible lever that bends that curve without destr quality. So this isn't optional. The question is execution.

Here's where it gets complicated. Deploying AI in a clinical environment is not l deploying AI in a SaaS product or a financial services workflow. The regulatory surface area alone is enormous. FDA has started asserting jurisdiction over clini decision support software in ways that are still being litigated in real time. HIPA creates data handling requirements that most generic infrastructure can't meet the box. And beyond compliance, there's the clinical validation problem, which i own category of hard. A model that performs well on training data and even on t data from an academic medical center can fail in surprising ways when it hits th actual workflow of a community hospital in rural Tennessee with different patie demographics, different EHR configurations, and different clinical protocols. Tł history of health tech is littered with products that worked in the demo and diec deployment.

Health systems know this. Ask any CIO at a regional health system what keeps t up at night and AI governance is going to be on the list, right next to cybersecur and Epic upgrade cycles. They're getting pitched constantly by AI vendors and t sophisticated enough to know that the pitch is usually ahead of the reality. The l ones will tell you they don't have the internal capability to evaluate AI tools rigorously, much less build the deployment infrastructure to operationalize ther safely. A mid-sized health system might have a dozen data scientists and maybe (two people with any real ML background. That's not enough to build model fine tuning pipelines, clinical validation frameworks, governance documentation, ch management playbooks, and the monitoring infrastructure to catch model drift (time. It's especially not enough to build all of that and then maintain it across d(of use cases simultaneously.

So you have a demand signal that is clear and growing, an internal capability gap is not going to close anytime soon, and a vendor landscape that is mostly offerin models rather than infrastructure. That's the setup.

The Mistake Everyone's Making: Bettin on Models

Walk into any health AI pitch and you'll hear some version of the same story. Th founders trained a model on clinical data, it outperforms GPT-4 on some bench that sounds important, and they're going to sell it to health systems that want be clinical decision support or revenue cycle automation or ambient documentatior model is the product. The moat is the model.

This is the wrong bet, and it's wrong for reasons that don't require a lot of hand wraving to see. Foundation model capability is improving so fast and being commoditized so quickly that any performance advantage a health-AI-specific m has today is likely to evaporate within 12-18 months as Anthropic, Google, and OpenAI continue pushing the frontier. These companies have research teams an compute budgets that no health tech startup can match. The gap between a gene purpose foundation model with good prompting and a health-specific fine-tunec model is narrowing every quarter. Betting a company on that gap holding is a fra thesis.

There's a second problem, which is that health systems are becoming increasing wary of model lock-in. Any sophisticated health system CIO who's thought abou for more than five minutes is worried about signing a long-term contract with a vendor whose core value proposition is a proprietary model, because they've wat the EHR market long enough to know what vendor lock-in looks like. The smart are already asking for model-agnostic architectures. They want to be able to swa foundation models as the market evolves without ripping out their deployment infrastructure.

The companies that are going to win in health AI over the next decade are the or that figure out how to make money on the deployment layer – the governance, th validation, the integration, the monitoring – rather than the model itself. This is new idea in enterprise software. The history of enterprise infrastructure is full c companies that didn't win by having the best underlying technology. They won t being the most trusted, most integrated, most operationally embedded layer in t stack. ServiceNow doesn't have better AI than Salesforce. Azure wasn't the best from a raw technical standpoint. They won by becoming the operational fabric t enterprises couldn't easily rip out, and by accumulating the operational data anc feedback loops that made their platforms better over time.

> Health AI infrastructure is that opportunity. And unlike foundation model competition, it's actually winnable.

What "Clinical Infrastructure" Actually Means

It's worth being specific here because "infrastructure" is one of those words that used to mean everything and therefore sometimes means nothing. In the context health AI, the clinical infrastructure layer has several distinct components that e represent real engineering and real value.

The first is model deployment and configuration management. Health systems d just need access to a model – they need a way to deploy it into their specific environment, configure it for their workflows, version it so they can roll back wh something goes wrong, and manage updates without breaking production systen Most health systems are running Epic or a combination of Epic and a few point solutions, and the integration surface is complex. There's HL7, there's FHIR, the lot of legacy data sitting in formats that require significant transformation befor model can do anything useful with it. The infrastructure company builds the connective tissue that makes all of this work without requiring the health systen maintain a team of integration engineers internally.

The second component is clinical validation pipelines. This is the piece that mos vendors don't want to talk about, because it's hard and it slows down sales cycle: it's also the piece that is most likely to cause catastrophic failure if it's missing. Clinical validation means establishing – rigorously, with actual clinical oversigh that a model performs as expected across the patient populations and workflow contexts it will encounter in production. It means building test sets that reflect t health system's actual demographics, not just the demographics of whoever contributed to the training data. It means defining the performance thresholds b which the model shouldn't be deployed, and having a process for evaluating moc behavior when it falls outside those thresholds. The infrastructure company bui the tooling that makes this process repeatable and scalable, rather than a bespok consulting engagement every time a new model or use case gets deployed.

The third component is governance and documentation. This is the piece that ke the compliance and legal teams from killing the deployment before it starts. FD Software as a Medical Device framework, if applied to clinical decision support, requires specific documentation about intended use, performance characteristic risk mitigation. Even for tools that don't rise to the level of regulated SaMD, hea system legal and compliance teams want documentation that demonstrates due diligence. Building the governance layer means creating the templates, the audit trails, the documentation workflows, and the review processes that make it poss deploy AI without a six-month compliance review every time. This is genuinely valuable and genuinely hard to build from scratch for each deployment.

The fourth component is monitoring and drift detection. Models degrade over ti Patient populations shift, coding practices change, clinical protocols evolve, and distribution of inputs the model sees in production starts diverging from what i in training. Without systematic monitoring, the health system has no idea wheth model that was performing well six months ago is still performing well today. Tl infrastructure layer includes the monitoring tools that track model performance production, alert when drift is detected, and provide the data needed to decide whether to retrain, reconfigure, or pull a model out of production.

The fifth component, and the one that's most underappreciated, is change management infrastructure. AI deployment failures in health systems aren't usu technical failures. They're adoption failures. Clinicians don't trust the tool, wor weren't redesigned to accommodate it, nobody trained the nursing staff, the aler firing at inconvenient moments, and eventually the thing gets turned off. The infrastructure company builds the playbooks, the training materials, the feedbac collection mechanisms, and the workflow integration guidance that makes adopt actually happen. This is part technology and part professional services, and it's t piece that most pure-technology companies don't want to do because it feels like consulting. But it's also the piece that drives adoption rates from 20% to 80%, an creates the kind of operational embeddedness that makes a customer relationshi genuinely sticky.

The Coalition Play: Network Effects in a Regulated Industry

Here's where it gets structurally interesting from an investment standpoint. The clinical infrastructure company described above is valuable as a standalone busi

But the version that becomes generationally defensible is the one that's built on coalition of health systems that share data, validation results, and deployment feedback across the network.

Think about what this looks like in practice. A coalition of, say, 30 health systen agrees to contribute de-identified clinical data and deployment feedback into a s infrastructure platform. The platform uses that pooled data to fine-tune and val models against a much more diverse set of patient populations and workflow cor than any single system could represent. When a new model gets deployed at mer system number 31, it's not starting from scratch – it's deploying into an infrastru that has already seen similar workflows at 30 other systems, already knows what common failure modes are, already has validation benchmarks calibrated agains comparable populations. The deployment goes faster, the validation is more rigo and the outcomes are better.

This creates a compounding network effect that is genuinely hard for point solu to replicate. More deployment sites mean more feedback data. More feedback da means better performance benchmarks and faster drift detection. Better benchn mean stronger safety validation that can be documented for regulators and healt system legal teams. Stronger safety validation means faster sales cycles and high willingness to pay at new member systems. The whole thing compounds in a way favors scale, which means the company that gets to meaningful coalition size firs a structural advantage that's difficult to overcome.

There's a historical analogy here worth taking seriously. Vizient, which is the lar healthcare group purchasing organization in the US with over $100B in annual purchasing volume, started as a coalition of academic medical centers that poole purchasing power to negotiate better supply chain contracts. It took decades to there, but the core insight was the same: health systems that compete with each on clinical outcomes are often willing to cooperate on operational infrastructure especially when the cooperation creates something none of them could build alo deployment infrastructure is an area where competition between health systems close to zero and the collective action problem is real. A coalition model is not ju nice-to-have for the business – it's the mechanism that creates the most durable

The data network effect in health AI is particularly powerful because the clinica validation challenge is population-specific in ways that make large, diverse train datasets worth a lot more than small, homogeneous ones. A model that's been validated across 30 health systems representing academic medical centers, comm hospitals, rural critical access hospitals, and safety net systems has a claim to generalizability that a model validated at one academic center simply can't make health system legal teams that are trying to defend AI deployment decisions to t boards and to regulators, that generalizability claim is genuinely valuable. It's th difference between "we tested this at Mayo Clinic" and "we tested this across 30 systems representing 3 million patient encounters annually."

The Business Model and Why It Works

There are two revenue streams here and both of them are attractive. The first is membership fees from coalition health systems. Coalition members get access tc shared infrastructure, the pooled validation data, and the deployment tooling, ar they pay an annual fee that scales with the size of their system and the number c use cases they're deploying. This is predictable, recurring revenue with high retention, because once a health system has built its AI deployment workflows o platform, switching costs are substantial. It's also a revenue model that aligns wi how health systems think about technology purchasing – they're used to paying subscriptions for operational infrastructure and the value proposition is straightforward.

The second revenue stream is licensing to non-member health systems. This is w the business really scales. Non-members can access the platform and the validat frameworks without contributing to the shared data pool – but they pay a highe for that access, because they're getting the benefit of the network without contributing to it. This is a common and effective pricing structure in coalition- platforms: members get a discount in exchange for data contribution, non-meml pay a premium. The non-member market is much larger than the coalition itself there are 900 meaningful health systems in the US, and a functional coalition mi have 30-50 of them. The other 850 are potential licensing customers.

At meaningful scale, the math works out to an interesting place. If the platform license to 200 health systems at an average contract value of $2.5M annually – w not an unreasonable number for enterprise infrastructure that's embedded in AI deployment workflows – that's $500M ARR from the non-member market alone. membership revenue and the number gets bigger. The comparable companies in adjacent markets – Veeva in life sciences, Evolent in value-based care, Health Ca in data analytics – are all trading or have traded at 8-12x ARR multiples for SaaS forward health IT businesses. The exit math at scale is compelling.

The cost structure is also attractive, especially relative to a model company. The marginal cost of adding a new deployment site to the platform is low once the co infrastructure is built. The company isn't paying Nvidia to run inference at scale that cost lives with the foundation model providers. The infrastructure company costs are primarily engineering (to maintain and extend the platform), clinical ar regulatory expertise (to maintain the governance frameworks and validation standards), and customer success (to drive adoption and expansion within memb systems). All three of those scale better than a services business, and the clinical regulatory expertise is something that compounds over time rather than having t rebuilt with each new customer.

What Could Kill This

It would be dishonest to write a 4,000-word pitch for this company without sper some time on what could go wrong. There are a few failure modes worth taking seriously.

The first is regulatory fragmentation. FDA's position on AI-powered clinical dec support is still evolving, and a major shift in how the agency classifies these tools could create compliance costs that slow deployment and make the governance la more expensive to maintain. This is a real risk, but it cuts in favor of the infrastructure company more than against it – tighter regulation increases the va standardized governance frameworks rather than decreasing it. If anything, a mo

aggressive FDA stance on clinical AI makes the company's regulatory infrastruct
more valuable, not less.

The second failure mode is EHR vendors eating the market. Epic in particular ha been aggressive about building AI capabilities natively into its platform, and it h data advantage of sitting at the center of more than half of US hospital IT environments. If Epic decides to become the clinical AI infrastructure layer rath than a point of integration, it would be a serious competitive threat. This is not a trivial concern. But there are reasons to think Epic won't fully eat this market. H systems that use Cerner or other EHRs are not going to build their AI infrastruc on Epic. And even health systems that use Epic are increasingly wary of further vendor concentration – they've watched what happens when you let one vendor control too much of the stack. There's a market for an independent, EHR-agnost infrastructure layer even in a world where Epic is building its own.

The third failure mode is coalition dynamics. Getting 20-30 health systems to ag data sharing governance, IP allocation, and cost-sharing arrangements is genuin hard. Health systems are not naturally cooperative and they have legal and comp teams that will find reasons to slow down any agreement that involves sharing da The company needs either exceptional relationship capital and dealmaking abili the founding team, or a compelling enough anchor member to create momentum pulls other systems in. Academic medical centers with strong technology transfe offices and reputational incentives to lead on AI governance are probably the be targets. If the company can get two or three AMCs as founding members, the sal pitch to community health systems becomes much easier.

The fourth risk is timing. Health system budgets are constrained right now – lab costs post-pandemic, Medicare reimbursement pressure, and capital expenditur overhang from facility investments are squeezing a lot of systems. Selling new infrastructure during a budget crunch is hard, even when the ROI case is clear. T counterargument is that AI deployment pressure is also coming from the board l at a lot of health systems, which creates a pathway to budget even in a constrain environment. Boards want to know their system has a credible AI strategy, and " joined the consortium that's building shared deployment infrastructure" is a cre answer that doesn't require massive internal investment.

The Investment Case

The venture case for this company is different from the typical health tech ventu case, and it's worth being explicit about that. This isn't a company that grows fa the early years and then scales. It's a company that takes three to five years to bu the coalition, establish the infrastructure, and accumulate the deployment feedb that makes the platform genuinely defensible. Series A and B are going to be ab proving that the coalition model works and that health systems will actually contribute data and pay for shared infrastructure. Series C and beyond are abou licensing scale and expansion into adjacent use cases.

That timing profile is actually an advantage from a competitive standpoint, beca selects against impatient capital and tourist investors who are going to lose inte when the company isn't at $100M ARR after three years. The right investor base people who understand that infrastructure businesses take time to build but pro extraordinary long-term returns when they do – the Benchmark, a16z bio, Gene Catalyst Health profiles rather than the growth equity crowd.

The founding team needs to be unusual. It needs clinical credibility, ideally som who has run AI or informatics at a health system and understands what the deployment problems actually look like from the inside. It needs regulatory sophistication – someone who has navigated FDA's digital health frameworks an understands how to build governance infrastructure that holds up to scrutiny. It needs enterprise sales capability, specifically health system sales, which is a gen different skill set from SaaS sales. And it needs someone who can build and mai the coalition relationships that are the foundation of the network effect. That's a ask of a founding team, which suggests this is likely a company that starts with a experienced team rather than a 25-year-old with a hot model.

The market timing, despite the budget pressures, is probably right. Health syste are being flooded with AI vendor pitches and they're starting to recognize that t need help sorting good from bad and deploying what they choose safely. Regulat are paying attention and are going to require more governance documentation o time. The foundation model market is mature enough that the deployment and governance layer is clearly the value-add, not the model itself. And there are eno health systems that have had bad early AI experiences – tools that got deployed without proper validation, promised workflows that never materialized – that th market is ready for an infrastructure company that does the hard work.

The AI clinical infrastructure company isn't the most exciting pitch in health te right now. It's not going to be on stage at JP Morgan talking about breakthrough model performance. It's going to be in the back room with the CIO and the CMI explaining how it solves the specific, boring, critical problem of making AI safe operational in their environment. That's exactly the kind of company that produ durable returns in enterprise software. The exciting model companies get acquir go public and then get competed into margin compression. The infrastructure companies become the fabric that nobody can remove, and they print cash for decades.

That's the bet.

5 Likes • 2 Restacks

← Previous

Nex

Discussion about this post

Comments Restacks

Write a comment...