The Picks and Shovels of Digital Healt Infrastructure Bets That Will Define t Next Decade

FEB 27, 2026 • PAID

Abstract

Digital health investment hit roughly $29B globally in 2023 after the correction the 2021 peak of $57B. Most of the capital that got torched in the downturn cha applications -- the shiny clinical tools, the consumer apps, the point solutions. W survived and what will win the next cycle is infrastructure. This essay makes the that the durable alpha in digital health sits one layer below the applications mos investors are looking at, in the unglamorous but defensible plumbing that every health tech builder will need. Key themes:

- Data interoperability and the persistence of fragmentation as a forcing functic middleware

- Identity and patient matching as an underappreciated category with moat characteristics

- Clinical AI infrastructure vs. AI applications -- why the former is the better be

- Provider revenue cycle as infrastructure, not software

- Compliance and regulatory infrastructure as a structural tailwind

- Market sizing: the infrastructure TAM that nobody talks about is north of $80 2030

Table of Contents

1. The Application Graveyard and What It Teaches Us

2. Data Infrastructure: The Fragmentation Dividend

3. Identity Is Infrastructure

4. The AI Stack Problem in Healthcare

5. Revenue Cycle as Structural Infrastructure

6. Compliance Infrastructure Is Having Its Moment

7. Where the Money Should Go

The Application Graveyard and What It Teaches Us

There is a reason that the 2020-2021 digital health bubble produced so many companies with impressive press releases and mediocre outcomes. The investme thesis for most of that capital was essentially: healthcare is broken, technology f broken things, therefore health tech. That logic is not wrong exactly, but it skips part where healthcare is not just broken, it is structurally different from every o industry that has been disrupted by software. The regulatory environment is uni hostile, the sales cycles are brutal, the procurement committees are designed to slowly, and the data that any application needs to work is scattered across thous of siloed systems in formats that were standardized in 1987. You can build a bea chronic care management product and spend three years failing to get EHR integration before you run out of runway. The graveyard is full of companies tha solved the clinical problem and could not solve the plumbing problem.

What this tells the sophisticated investor is something that should feel obvious i retrospect: the constraint in health tech is almost never the application layer. Th are plenty of smart clinician-founders who understand the workflow problem. T

constraint is everything underneath. The data access problem. The identity prob
The compliance problem. The billing and remittance problem. The AI model evaluation problem. These are unsexy, technically hard, and -- crucially -- sharec across every application that anyone is trying to build. That combination of characteristics is exactly what makes them good businesses. Picks and shovels du a gold rush tend to be better investments than individual miners, and healthcare the early innings of a multi-decade software transformation that is going to requ lot of shovels.

The market correction that ran from late 2021 through 2023 was healthy. It rationalized valuations, it killed zombie point solutions that were never going to achieve the unit economics required for standalone success, and it concentrated capital toward companies that had real infrastructure value. What has emerged f that correction is a cleaner view of where durable value lives. The companies tha their valuations through the downturn were disproportionately infrastructure pl The companies that got torched were disproportionately application-layer point solutions dependent on grant funding, employer pilots, or consumer acquisition economics that never penciled.

The forward thesis, then, is not complicated: the next cohort of successful digita health applications is going to be built on a layer of infrastructure that mostly d not exist yet in mature form. The entrepreneurs building that infrastructure are to have better unit economics, stronger moats, and cleaner exit paths than the application builders they serve. The investors who figure this out early are going make a lot of money. The ones who keep chasing the clinical application story ar going to keep getting frustrated by the same structural problems that killed the cycle's darlings.

Data Infrastructure: The Fragmentatio Dividend

Anyone who has spent more than a week trying to build a data product in health has encountered what might be described as the fundamental absurdity of the sp

There are roughly 900 EHR vendors in the US market. The top three -- Epic, Ora Health (formerly Cerner), and Meditech -- account for maybe 60% of hospital be but the long tail is vast and weird and deeply entrenched. On top of that you hav claims ecosystem, which runs through clearinghouses and payers and generates in X12 EDI formats that look like they were designed to be unreadable. Then th pharmacy data, lab data, ADT feeds, device data from wearables and monitoring equipment, social determinants data, and an increasingly chaotic landscape of pa generated data from apps that do not talk to anything. The fact that any clinical decision support product works at all is kind of a miracle.

The federal government has been trying to solve this for about 15 years. The HIT Act in 2009 spent roughly $38B in Medicare and Medicaid incentives to get prov onto EHRs, which worked in the sense that EHR adoption went from around 12% hospitals in 2009 to over 96% today. What it did not solve was interoperability, because getting data into structured electronic systems and getting data to flow between systems are completely different problems. The 21st Century Cures Act the associated ONC information blocking rules that went into effect in 2022 mo the needle somewhat -- they created FHIR API mandates that made it meaningfu easier to pull patient data with patient consent -- but they did not solve the ente data problem for providers who want to do population health analytics or AI mo training.

The infrastructure play here is what could be called the fragmentation dividend. Because healthcare data is so fragmented, the companies that can aggregate, normalize, and deliver it have structural leverage over every application built on The most defensible version of this business is a network model -- the more paye providers, labs, and pharmacies connected to your network, the more useful you product is, which attracts more endpoints, which makes the network more valua This is a flywheel that is genuinely hard to replicate once it reaches critical mass Health data network companies that have achieved multi-modal connectivity acr claims, clinical, and pharmacy data are not easy to displace. Their switching cost high, their data coverage creates real differentiation, and the compliance

requirements around data handling (HIPAA, state privacy laws, emerging AI governance requirements) make it harder, not easier, for new entrants to compet

The specific opportunity right now is in the normalization and enrichment layer between raw data aggregation and application consumption. Raw FHIR data is technically standardized but practically inconsistent -- problem lists use differe coding conventions, medication records have reconciliation issues, lab values co without proper reference ranges. The companies building the normalization and clinical NLP infrastructure to make raw data actually usable are solving a proble that every AI company in healthcare is going to hit. The total addressable marke data infrastructure in healthcare has been estimated at north of $15B by 2028, a that number is almost certainly conservative given the acceleration of AI use cas that all require clean training and inference data.

Identity Is Infrastructure

Here is a fact that should be alarming to anyone thinking about healthcare AI at there is no national patient identifier in the United States. Congress actually prohibited HHS from establishing one back in 1998, and while there have been periodic efforts to revisit that decision, the political dynamics around privacy ha made it a non-starter. The practical consequence is that patient matching -- the process of correctly linking records for the same patient across different systems done probabilistically, using algorithms that match on combinations of name, da birth, address, and other demographic fields. Error rates for probabilistic patien matching range from around 7% to 20% depending on the system and patient population. For a country spending $4.5T annually on healthcare, that is a stagg amount of misidentification.

The implications compound quickly. Duplicate records create clinical safety issu if a clinician cannot see a complete medication list because half the patient's rec are under a slightly different name, that is a patient safety problem. They create billing problems -- matching remittance to claims requires accurate patient iden and mismatches create write-offs and rework. They create research problems --

longitudinal patient data analysis requires accurate identity linkage across time across institutions. And in the context of AI model development, they create trai data problems -- models trained on patient records contaminated by identity err will learn spurious correlations.

The companies building enterprise master patient index (EMPI) infrastructure a probabilistic matching algorithms have been around for a while, but the space is getting interesting again for a couple of reasons. First, the explosion of new data -- wearable data, social determinants data, digital biomarkers -- is adding new dimensions to identity that traditional EMPI systems were not built to handle. Second, the AI governance requirements emerging from the FDA's Digital Healt Center of Excellence and from state-level AI bills are going to require better dat provenance, which requires better identity infrastructure. Third, the shift towar value-based care creates attribution requirements -- you have to know which pat belong to which care team to do population health, and that requires reliable ide

The moat characteristics here are strong. Good EMPI infrastructure requires lar reference datasets of known identity linkages to train matching algorithms, and datasets take years to accumulate. There are network effects -- identity confiden improves as more systems contribute identity data to a shared graph. And the compliance requirements around PHI handling create a meaningful barrier for r entrants who have not already built the HIPAA-compliant data handling infrastructure. It is not a flashy category, but an investor who can identify the EI infrastructure company that is going to be the identity backbone for value-based contracting and AI model governance is going to do very well.

The AI Stack Problem in Healthcare

Every health system CEO in America is currently telling their board that they ar pursuing an AI strategy. Most of them are not wrong, exactly, but what they are actually doing is considerably more complicated than deploying GPT-4 on their patient portal. Clinical AI has a set of requirements that general-purpose AI

infrastructure was not designed to handle, and the companies figuring out how t bridge that gap are building businesses that will compound for a long time.

Start with the data problem, which was covered above but bears repeating in the context. Training a clinical AI model requires labeled clinical data. Getting labe clinical data requires either annotating it yourself (expensive, slow) or buying ac to de-identified datasets (limited, potentially biased). The de-identification prob itself is non-trivial -- safe harbor de-identification under HIPAA requires remov specific data elements, but it also requires that the data holder not have actual knowledge that the remaining information could identify an individual. For rare diseases or small geographic areas, de-identified data can still be re-identifiable. companies building compliant synthetic data generation and privacy-preserving federated learning infrastructure are solving a problem that every clinical AI developer has, and the market for clinical AI model development infrastructure going to be enormous.

Then there is the model validation problem. The FDA has cleared over 500 AI/M based Software as a Medical Device products as of mid-2024, but the regulatory pathway is still evolving. The predetermined change control plan framework tha FDA introduced in 2021 gave developers a way to update models post-approval without seeking new clearance for every change, but it requires robust monitori infrastructure to demonstrate that model performance is not degrading in deployment. Health systems deploying clinical AI are going to need tooling for continuous model performance monitoring, drift detection, and bias auditing ac patient subpopulations. That tooling does not really exist at scale yet, and the companies building it are going to be critical infrastructure for the AI governan requirements that are coming.

The third layer of the AI stack problem is inference infrastructure optimized fo clinical workflows. General-purpose LLM APIs are not designed for the latency reliability requirements of clinical decision support tools deployed at the point c care. A hospitalist using an AI-assisted differential diagnosis tool cannot wait th seconds for a response, and the system cannot be down during a night shift beca API provider is having scaling issues. The companies building healthcare-specif

inference infrastructure -- optimized for clinical terminology, designed for high availability deployment, with the audit logging required for clinical AI governan are building something that general-purpose cloud providers are not going to prioritize. Epic and Oracle Health will eventually build more of this natively, bu multi-EHR interoperability requirement means there will always be space for independent clinical AI infrastructure.

Practically speaking, the infrastructure bet in AI is not on any specific clinical A application. It is on the model operations platform, the synthetic data companie federated learning infrastructure, and the AI governance tooling that every clini developer is going to need. The market sizing here is genuinely hard to pin dow because the category is so new, but if even 10% of the $130B+ that US health sys spend annually on IT shifts toward AI, the infrastructure supporting that AI wil capture a meaningful fraction of that spend.

Revenue Cycle as Structural Infrastructure

Revenue cycle management is one of those categories that investors sometimes underestimate because it sounds like back-office software. It is not. Revenue cyc the financial operating system of every provider organization in the country, and complexity of the US claims-based reimbursement system means that revenue cy infrastructure has structural characteristics that make it extremely durable as a business.

To understand the opportunity, it helps to understand the dysfunction. A typica health system writes off somewhere between 2% and 5% of net patient revenue t debt and underpayments. For a $2B revenue health system, that is $40M to $100 walking out the door annually. The sources of leakage are numerous: prior authorization denials that were not appealed, coding errors that result in downc claims, eligibility verification failures that result in uncompensated care, coordi of benefits issues that result in underpayment from primary payers, and release information workflows that are manual enough to create compliance exposure. N

of these problems are new, but they are getting worse as payer behavior has becc more aggressive. Prior authorization denial rates have increased substantially -- AMA's 2023 prior authorization survey found that 93% of physicians reported ca delays due to prior authorization requirements, and 25% reported that prior authorization had led to a serious adverse event for a patient.

The infrastructure play in RCM is not in the traditional billing company model, which is mostly a services business with modest technology leverage. It is in the automation infrastructure that sits inside the RCM workflow. Prior authorizatio automation, powered by AI that can predict which procedures are likely to get d and pre-populate the clinical justification, is a category that is moving fast. Deni management automation, which uses ML to triage denials by appeal probability pre-populate appeal letters with relevant clinical documentation, is another. The companies building this automation infrastructure and selling it as a platform to health systems, billing companies, and specialty practices are compounding on t a base business that is not going away -- the US is not moving away from fee-for- service billing in any meaningful timeframe, and the complexity of multi-payer I is only going to increase as value-based contracts layer additional reconciliation requirements on top of traditional claims.

There is also an interesting intersection between RCM infrastructure and releas information (ROI) infrastructure that is worth flagging. ROI is the process by wh health systems respond to requests for patient records from payers, attorneys, pa themselves, and other providers. It is heavily regulated, operationally intensive, a generates substantial revenue for the health systems that manage it well. The automation of ROI workflows -- using AI to classify requests, route them to the appropriate handler, apply the correct redaction logic for different request types track response time compliance -- is a category that has historically been served large outsourced services companies. The shift toward software-native ROI infrastructure that captures the margin currently going to outsourcers is a comp play, particularly as the health system market continues consolidating into syste large enough to internalize these workflows.

Compliance Infrastructure Is Having Its Moment

The regulatory environment for digital health is getting more complicated, not l The 21st Century Cures Act information blocking provisions, the ONC HTI-1 ru published in 2024 that extended FHIR interoperability requirements to new dat classes, state privacy laws in California, Washington, and others that layer addit requirements on top of HIPAA, the FTC's increasing scrutiny of health data pra -- all of this creates demand for compliance infrastructure that was not necessar years ago.

The specific pressure points are worth enumerating because they each represent market. HIPAA has been the baseline for health data privacy for 25 years, but enforcement is getting more aggressive. OCR settlements have increased in freq and magnitude -- 2023 saw several multimillion-dollar settlements related to pat portal data sharing practices and tracking pixel use. The Wave of lawsuits follov the revelation that common web analytics tools were transmitting PHI to Meta a Google created a new category of risk that most health systems had not adequate assessed. The companies building privacy engineering infrastructure -- tools for automated PHI detection in web and mobile applications, consent management platforms designed for the complexities of health data, and breach risk assessme tooling -- are seeing tailwinds from every enforcement action and settlement.

The information blocking provisions of the 21st Century Cures Act are creating different kind of compliance opportunity. Providers, health IT developers, and h information networks are now prohibited from engaging in practices that unreasonably interfere with the access, exchange, or use of electronic health information, subject to eight defined exceptions. Understanding which data shar practices are protected by those exceptions and which are not is genuinely complicated, and most health systems do not have internal counsel with the dep stay current on ONC guidance. The companies building compliance tooling specifically for information blocking risk assessment and documentation are fill real gap.

State AI legislation is the most forward-looking compliance infrastructure opportunity. Colorado passed SB21-169 in 2021 requiring insurers to test algori for bias before deployment, and several states are now working on broader healt governance requirements. The FDA is developing its regulatory framework for clinical decision support AI, and the Biden administration's executive order on A created additional reporting requirements for high-impact AI systems including used in healthcare. The companies building AI governance infrastructure -- mod documentation, bias testing, audit trails, and regulatory submission support -- a positioned at the intersection of a technology wave and a regulatory wave that ar going to collide hard in the next few years.

The compliance infrastructure market is unusual in one respect that makes it particularly attractive for investors: demand is driven by regulatory requirement rather than discretionary IT budget. Health systems cannot not comply with HIP cannot not respond to information blocking complaints, and cannot not address governance requirements as they solidify into law. That creates a revenue base th more predictable and less economically sensitive than application-layer health IT spend. It also creates sales cycles that, while still long by normal software standa have a forcing function that pure application software does not.

Where the Money Should Go

Pulling this together into an investment framework requires being honest about things. Infrastructure is not a monolithic category -- there is a meaningful differ between the dynamics of a data network business, an AI governance platform, an EMPI company. Each has different capital requirements, different sales cycles, a different exit dynamics. But they share a common characteristic that distinguish them from the application layer: they are shared services for the ecosystem rathe point solutions for specific workflows, which means their revenue is diversified a many buyers and their value accretes as the ecosystem grows.

The data network businesses are the most capital-intensive and have the longest to defensibility, but they also have the highest ceiling. A company that achieves

connectivity across 5,000 providers, 300 payers, and the major pharmacy chains i sitting on an asset that is very difficult to replicate and that every AI company in healthcare wants access to. The challenge is getting to that critical mass, which requires patient capital and a management team that understands network busir The right questions for diligence are about the rate of endpoint addition, the de data coverage at connected sites, and the robustness of the normalization layer. Shallow connectivity that pulls only FHIR appointment data is not the same as d connectivity that includes ADT feeds, claims data, and medication management.

The AI infrastructure stack businesses are more varied. Synthetic data companie federated learning platforms are pre-product-market fit for most buyers right n health system AI teams are sophisticated enough to understand the need but procurement processes for new infrastructure categories take time. The model operations and governance platforms are closer to the money because they are sc a problem that is already creating liability for health systems deploying AI. An investor who gets into the right AI governance platform at the Series A is buying a category that will be table stakes within five years. The comps are companies li Veeva in life sciences compliance or Egnyte in healthcare document managemen infrastructure categories that seemed niche until regulatory pressure made them mandatory.

Revenue cycle automation infrastructure is probably the most immediately monetizable category. The ROI on prior authorization automation is measurable weeks, not years -- if an AI system reduces prior auth denial rates by 15% for a h system with $500M in annual prior auth-related revenue exposure, the math is straightforward. The sales cycle is still multi-quarter, but procurement decision be justified with clear financial modeling. The companies to watch here are the c building workflow automation that integrates natively into existing EHR and pra management systems rather than requiring rip-and-replace implementations. Th biggest risk in this category is EHR vendor encroachment -- Epic has been stead building more revenue cycle functionality natively, and any infrastructure play t depends on Epic workflow gaps is making a bet against Epic's product roadmap.

Compliance infrastructure is the most policy-sensitive category, which means it carries regulatory risk alongside regulatory tailwind. A company building exclus for information blocking compliance is vulnerable to changes in ONC enforcem posture. The better bet is companies building horizontal compliance infrastruct that serves multiple regulatory requirements -- HIPAA, information blocking, st governance, FDA AI oversight -- rather than point solutions for a single require The platform plays in compliance infrastructure should be able to demonstrate t they can extend to new requirements as they emerge without requiring product rebuilds.

The through-line across all of these categories is that the health tech infrastruct market is being underserved relative to its strategic importance. Venture capital flows disproportionately toward clinical applications because the storytelling is -- a chronic care management platform is easier to explain to a generalist investo than a FHIR normalization engine. But the returns to infrastructure investing in healthcare are going to be disproportionately strong over the next decade, for th same reason that AWS returns were disproportionate to the returns on web applications in the 2010s. The picks and shovels won the first digital health cycl investors just were not paying attention to the right layer.

The founders building in these categories tend to be former health IT operators have been close enough to the plumbing to understand how bad it actually is. Th diligence signal is a founding team that can describe the problem with the speci granularity of someone who has tried and failed to solve it using existing tools. A founder who can explain exactly why HL7 v2 ADT feeds are structurally differen from FHIR R4 resources and why that matters for a specific use case is more cre than one who has learned the talking points from a Gartner report. The domain specificity required to build well in this space is a feature, not a bug -- it is part what keeps competition manageable and preserves margin.

The market sizing for digital health infrastructure broadly, encompassing data networks, identity, AI infrastructure, RCM automation, and compliance tooling, north of $80B by 2030 by most reasonable estimates, and that number does not f account for the AI-driven expansion of what infrastructure needs to do. The

companies winning these categories are going to be worth substantially more tha most current health tech unicorns because their revenue is stickier, their growth tied to the expansion of the entire ecosystem rather than penetration of a single market, and their exit multiples reflect infrastructure economics rather than Saa point solution economics. That is where the sophisticated capital in digital healt should be going.

3 Likes • 2 Restacks

← Previous

Nex

Discussion about this post

Comments Restacks

Write a comment...