

The Hidden Economics of Truth: Why Data Curation Will Make or Break Healthcare AI

OCT 31, 2025



Share

Disclaimer: The thoughts and opinions expressed in this essay are my own and do not those of my employer.

Abstract

Healthcare artificial intelligence has reached an inflection point where the bottleneck is no longer model architecture or compute power, but the quality and curation of training data. This essay examines the economic structures underlying data curation in healthcare AI, exploring why traditional approaches to data acquisition and labeling are failing, how the unit economics of curation create asymmetric advantages, and what this means for the future competitive landscape of healthcare AI companies. Key topics include the true cost of expert annotation, the emergence of curation as a competitive moat, the role of synthetic data generation, and the structural reasons why large language models cannot solve healthcare's data quality problem alone. For health tech entrepreneurs and investors, understanding these economics is critical to evaluating which AI healthcare companies will achieve sustainable margins and which will struggle with permanently deteriorating unit economics as they scale.

Table of Contents

The Curation Crisis Nobody Talks About

The Unit Economics of Medical Truth

Why Expert Networks Are the New Moat

The Synthetic Data Mirage

Feedback Loops and Compounding Advantages

What This Means for Healthcare AI Investment

The Path Forward

The Curation Crisis Nobody Talks About

Every healthcare AI pitch deck includes a slide about their proprietary dataset. Millions of medical images, billions of claims records, decades of electronic health record data scraped from health systems eager to monetize their digital exhaust. The implication is always the same: more data equals better models equals inevitable success. But this narrative fundamentally misunderstands the economics of what actually makes healthcare AI work at scale. The real constraint is not data volume but data curation, and the unit economics of curation in healthcare are unlike anything we have seen in consumer AI.

Consider what happened to most of the well-funded radiology AI companies from 2017 to 2020. They raised substantial Series A and B rounds based on impressive accuracy metrics derived from large datasets of medical images. The datasets were real, the accuracy improvements were measurable, and the potential market was enormous. Yet many of these companies have quietly pivoted, been acquired for modest multiples, or simply disappeared. The post-mortems rarely mention it explicitly, but dig into the operational details and you find the same pattern: the cost of maintaining and improving their models at scale exceeded what the market would bear. They drowned not in a lack of data, but in the escalating costs of curating that data to a standard that clinicians would actually trust and regulators would approve.

The economics are brutal and non-obvious. A single medical image might require thirty minutes of a board-certified radiologist's time to properly annotate for edge cases and rare pathologies. At three hundred dollars per hour for expert review,

one hundred and fifty dollars per image for high-quality ground truth. Scale that across the ten thousand examples you need for a robust model in a specific subdomain, and you are looking at one point five million dollars just for initial training data curation. But the real cost explosion happens post-deployment when you discover that model performance degrades in the wild, new edge cases emerge, and you need continuous re-curation to maintain clinical utility. Suddenly you are spending two hundred thousand dollars per quarter on ongoing data quality, and cost scales linearly with every new clinical context you try to address.

Compare this to consumer AI where you can use crowdsourced labels, automated quality checks, and user feedback loops to curate data at a fraction of the cost. Reddit moderators and Mechanical Turk workers can label images for dollars, not hundreds of dollars. The accuracy threshold is also completely different. If a consumer recommendation algorithm is eighty-five percent accurate, users shrug and scroll past the bad recommendations. If a diagnostic algorithm is eighty-five percent accurate, people die and lawyers get involved. Healthcare AI requires not just high accuracy but also explainability, auditability, and liability-grade certainty about why the model made each decision. All of that traces back to curation quality.

This curation crisis creates a perverse situation where the companies with the most data are not necessarily the ones with the best models. In fact, having access to massive uncured datasets can be a liability rather than an asset because it creates the illusion of readiness while hiding the true cost of achieving clinical-grade performance. I have watched multiple startups burn through tens of millions in funding trying to clean and curate datasets that were fundamentally too noisy to salvage. They would have been better off starting with ten thousand meticulously curated examples than ten million messy ones.

The Unit Economics of Medical Truth

To understand why healthcare AI curation is so expensive, you need to understand the labor market for medical expertise. Board-certified physicians in the United States earn between two hundred thousand and five hundred thousand dollars annually

depending on specialty, and their time is the scarcest resource in healthcare. When you ask a radiologist to spend thirty minutes carefully annotating an X-ray with bounding boxes, segmentation masks, and differential diagnoses, you are competing with the three hundred dollars they could earn reading clinical cases in that same time period. The opportunity cost is real and it sets a floor on curation costs that the amount of scale economies can break through.

But the economics get worse when you account for the expertise curve. Not all physicians are equally good at annotation. A first-year radiology resident will miss subtle findings that a fellowship-trained subspecialist catches immediately. For healthcare AI applications, you do not just need any doctor, you need doctors in the ninety-fifth percentile of their specialty who can catch the edge cases and rare presentations that will make or break your model in clinical practice. These physicians are even more expensive and even harder to recruit for annotation work because they have the most lucrative clinical opportunities.

This creates a fundamental scaling problem. As you try to expand your healthcare model to cover more conditions, more patient populations, and more clinical conditions, your curation costs do not decrease, they increase. Each new domain requires a new set of expert annotators with different subspecialty training. Your cost per curation training example stays constant or rises even as you scale, which is the opposite of what venture capitalists expect from software businesses. Traditional software has beautiful unit economics where marginal costs approach zero as you scale. Healthcare AI curation has flat or rising marginal costs, which fundamentally changes the investment math.

Some companies have tried to address this through hierarchical annotation workflows where less expensive annotators do initial labeling and experts only review edge cases. This can reduce costs by perhaps fifty percent but it introduces new quality control problems. How do you know which cases are edge cases that need expert review and which are straightforward cases that junior annotators can handle? You essentially need an initial model to triage the annotation work, which creates a bootstrapping problem. Your initial model is not good enough to accurately identify edge cases, so you end up

either over-triaging to experts at high cost or under-triaging and missing important cases that degrade model performance.

The other approach companies have tried is training clinicians to be better annotators through detailed protocols and feedback. This works to some extent but it runs into the fundamental problem that clinical judgment is tacit knowledge that resists codification. A radiologist knows a mass is suspicious based on pattern recognition developed over thousands of cases, but they cannot always articulate the precise features that trigger their suspicion in a way that translates to annotation guidelines. The things that make someone a good clinician, intuition and gestalt reasoning, precisely the things that make it hard to standardize their annotation work.

There is also a hidden cost in annotation latency. When you need expert review, you cannot just spin up annotation capacity on demand the way you can with crowdsourced labeling. Physicians have clinical schedules, administrative burdens and limited availability. It might take weeks or months to get a batch of complex cases reviewed by the right subspecialists. This latency kills your ability to iterate quickly on model improvements, which means you cannot use the rapid experimentation that makes modern AI development efficient. You are stuck in a world where each model iteration requires months of curation work, which dramatically slows you down relative to competitors in other domains.

Why Expert Networks Are the New Moat

Given these economics, the companies that will win in healthcare AI are not necessarily those with the most data or the best ML engineers. They are the companies that have built sustainable, scalable pipelines for accessing medical expertise at reasonable cost. This is leading to a fundamental shift in what constitutes a defensible moat in healthcare AI. Five years ago, the moat was proprietary data access through hospital partnerships. Today, the moat is expert network effects and annotation infrastructure.

The smartest healthcare AI companies are building what amount to internal expert networks, panels of subspecialist physicians who they can task for annotation work.

an ongoing basis. These are not one-off consulting arrangements but long-term relationships where physicians become deeply familiar with the company's annotation protocols, quality standards, and clinical use cases. Over time, these expert annotators become dramatically more efficient because they understand exactly what the model needs and how to provide feedback that improves performance. A radiologist who has annotated five thousand chest X-rays for your specific model can probably annotate new cases three times faster than a new annotator while maintaining higher quality.

This creates a powerful moat because your expert network becomes increasingly valuable over time and is nearly impossible for competitors to replicate quickly. Competitors cannot just hire a new panel of physicians and expect them to immediately match the efficiency and quality of your existing network. There is substantial institutional knowledge embedded in how your annotators work with your data that takes years to develop. This is similar to how the best consumer internet companies built operational advantages through things like supply chain relationships and community management that were not obvious from the outside but created sustainable competitive advantages.

Some companies are taking this even further by making expert annotators into stakeholders through equity compensation or revenue sharing arrangements. If you can align physicians financially with your company's success, you create retention that goes beyond just paying market rates for their time. A radiologist who owns equity in your company has an incentive to provide consistently high-quality annotations and to recruit other talented annotators to your network. This is particularly powerful in subspecialties where physicians know each other and reputation matters. One highly respected neuroradiologist can bring in five or ten colleagues, giving you rapid access to expertise that might otherwise take years to build.

The expert network moat also compounds with model improvements. As your model gets better, it can take on more of the easy cases, which means your expert annotators can focus on the truly challenging cases where their judgment adds the most value. This improves the efficiency of your curation spend because you are directing expensive expertise toward the places where it matters most. Your experts also become better at understanding what kinds of examples will most improve your

model, so they can proactively surface interesting cases rather than just reactively annotating whatever you send them. This virtuous cycle is hard to replicate and stronger over time.

There is also an interesting dynamic where expert networks create better product market fit. Physicians who annotate your training data often become early adopters and advocates for your product because they understand its capabilities and limitations better than anyone. They have seen thousands of examples of what the model gets right and wrong, so they know exactly how to use it effectively in clinical practice. These physician annotators become your best salespeople and clinical champions, which is invaluable in healthcare where trust and peer recommendations drive adoption more than traditional marketing.

The Synthetic Data Mirage

One of the most seductive narratives in healthcare AI right now is that synthetic data will solve the curation cost problem. The pitch goes like this: instead of spending millions on expert annotation, we will use generative AI to create unlimited synthetic medical data that looks realistic enough to train models. We can generate thousands of synthetic chest X-rays with perfectly labeled pathologies, synthetic EHR data with known ground truth diagnoses, and synthetic clinical notes with annotated entities and relationships. This will let us train models at a fraction of the cost while avoiding privacy concerns around real patient data.

This narrative is appealing but mostly wrong for reasons that become obvious once you understand the economics. The fundamental problem is that synthetic data only helps if the generator itself is trained on high-quality curated data. A generative model for medical images needs to be trained on thousands of expertly annotated medical images to learn what pathologies look like and how they vary across patients. You cannot escape the curation cost, you just move it upstream into training the generator. And because medical synthetic data requires even higher fidelity than consumer synthetic data, you actually need more expert curation to validate that the generator is producing clinically realistic examples.

There is also a subtle but important problem with distribution shift. Synthetic data generators learn the distribution of their training data, but real clinical practice includes all kinds of edge cases, rare presentations, and equipment artifacts that are hard to capture in synthetic generation. If your model is trained primarily on synthetic data, it will perform well on cases that look like typical textbook examples but fail on the weird edge cases that define real clinical utility. You end up needing substantial real-world data anyway to handle these edge cases, which means synthetic data does not actually eliminate your curation costs, it just changes when you incur them.

The more promising use case for synthetic data is not replacing expert curation but augmenting it for specific purposes. If you have a small set of expertly curated examples of a rare pathology, you can use synthetic data generation to create variations that help your model learn robustness to imaging conditions, patient positioning, and equipment differences. This is valuable but it is a complement to expert curation, not a replacement. You still need the initial expert-curated examples to ensure your synthetic variations are clinically realistic.

There is also an emerging concern about model collapse when synthetic data is used too aggressively in training. Recent research has shown that models trained on synthetic data generated by other models can develop increasingly narrow representations that miss important features of the real data distribution. In consumer AI this might just mean your chatbot sounds repetitive. In healthcare it could mean your diagnostic model systematically misses certain presentations of disease. The stakes are too high to rely heavily on synthetic data without extensive validation using expert-curated real-world examples.

The one area where synthetic data economics might actually work is in creating adversarial examples for robustness testing. Once you have a well-trained model, you can use synthetic data generation to create challenging edge cases that probe the model's decision boundaries. This is much cheaper than finding rare real-world examples and it helps you identify failure modes before deployment. But even here, you need expert clinicians to review the synthetic adversarial examples and confirm their clinical relevance.

they are clinically meaningful rather than just statistical outliers. The expert curation requirement never fully goes away.

Feedback Loops and Compounding Advantages

The companies that understand curation economics are building something more sophisticated than just expert networks. They are building closed-loop systems where deployed models generate feedback that improves curation efficiency over time. This is where the economics start to look more like traditional software and less like services business, but it requires careful design to achieve.

The basic idea is to instrument your deployed model to capture cases where it is uncertain or where clinicians override its recommendations. These cases are automatically flagged for expert review and added to your training data pipeline. Over time, this creates a flywheel where more usage generates more high-value training examples, which improves the model, which drives more usage. The key is that you are selectively curating the examples that matter most rather than trying to curate everything, which dramatically improves the efficiency of your curation spend.

This feedback loop approach requires solving several technical and operational challenges. You need infrastructure to capture the right signals from clinical usage, a system to prioritize which cases get expert review, and workflows that make it easy for clinicians to provide feedback without disrupting their work. Most healthcare AI companies underinvest in this infrastructure because it is not as exciting as model development, but it is often the difference between a product that gets better over time and one that stagnates after initial deployment.

The economic impact of these feedback loops is substantial. If you can reduce your cost of identifying high-value training examples by even fifty percent through automated flagging, you effectively double the efficiency of your expert network. Unlike the one-time cost reduction from process improvements, feedback loops compound over time. Each model improvement makes the automated flagging better, which makes expert review more efficient, which enables more model improvement.

This compounding effect is what eventually gives you software-like economics where marginal costs decline as you scale.

There is also a strategic dimension where feedback loops create switching costs for customers. A healthcare system that has been using your model for two years has inadvertently helped you curate thousands of examples specific to their patient population, equipment, and clinical workflows. Your model is now better for them than for anyone else, which makes it harder for competitors to displace you even if they have a technically superior model. The value is not just in the algorithm but in the accumulated institutional knowledge embedded in your curated dataset.

The best feedback loop designs also create network effects across customers. If you can aggregate learnings from multiple healthcare systems in a privacy-preserving way, each new customer makes your model better for all existing customers. A rare presentation that shows up at one hospital gets added to your training data and improves detection across your entire network. This is how you get true platform economics in healthcare AI rather than just selling point solutions. But it requires sophisticated data governance and privacy infrastructure that most early-stage companies lack.

What This Means for Healthcare AI Investment

For investors evaluating healthcare AI companies, the curation economics framework suggests a different set of diligence questions than the typical AI investment checklist. Yes, you should still care about model performance metrics and technical team quality. But the questions that really matter are about the sustainability and scalability of the company's approach to data curation.

How does the company source expert annotations and what is their fully loaded cost per curated training example? Many companies will quote you the hourly rate they pay annotators without accounting for overhead, quality control, rejected annotations, and the management time required to coordinate expert networks. Get to the true unit economics including all these costs. Then project how those costs scale as the

company tries to expand to new clinical domains or patient populations. If costs linearly or super-linearly with expansion, you are looking at a services business disguised as a software company.

What percentage of the company's ongoing operating costs go to data curation v model development and product engineering? Early-stage companies might spend thirty to forty percent of their budget on curation as they build initial training datasets. But if a growth-stage company is still spending this much on curation, suggests they have not built sustainable feedback loops and their unit economics not improve as they scale. You want to see curation as a percentage of operating declining over time as automation and feedback loops take over more of the work.

Does the company have a formal expert network or are they hiring annotation workers project by project? Project-based annotation is much more expensive in the long run because you lose institutional knowledge and efficiency gains between projects. Companies with formal expert networks, especially those with equity incentives for key annotators, have a structural advantage that compounds over time. Ask about annotator retention rates and tenure. If the company is churning through annotators every six months, their curation efficiency will never improve.

How sophisticated is their feedback loop infrastructure? Can they automatically identify high-value training examples from production usage? Do they have workflows that make it easy for clinicians to provide feedback without disrupting their work? Many companies claim to have feedback loops but when you dig into the details, it's just a ticket system where product managers manually review customer complaints and occasionally add cases to the training set. Real feedback loops are automated, continuous, and built into the product from day one.

What is the company's strategy for handling edge cases and distribution shift? Healthcare is full of situations where models trained on one population fail on another because of differences in disease prevalence, imaging protocols, or clinical workflows. Companies that have thought carefully about curation economics usually have specific plans for how they will identify and address distribution shift, often involving ongoing partnerships with diverse healthcare systems that provide

continuous feedback. Companies that have not thought about this will often handwave about fine-tuning and transfer learning without acknowledging the curation costs these approaches require.

The investment returns also look different when you account for curation economics. Healthcare AI companies are unlikely to have the winner-take-all dynamics of consumer internet because the curation advantages are not as winner-take-all as network effects in social media or marketplaces. Multiple companies can build successful expert networks in the same clinical domain and compete effectively. This suggests a market structure with a few dominant players per subdomain rather than one global winner. It also means you should be more conservative in projecting long-term margins because the curation costs create a permanent drag on profitability that software businesses do not face.

There is one caveat to this framework which is that companies building horizontal infrastructure rather than vertical clinical solutions face different economics. If you are selling tools that help other companies curate their own medical data more efficiently, your unit economics can look much more like traditional software. The challenge is that these infrastructure businesses need to reach substantial scale to capture significant value, and they face competition from both big tech companies and open-source alternatives. But for investors who believe curation is a fundamental bottleneck, infrastructure companies that meaningfully reduce curation costs could be very large outcomes.

The Path Forward

The curation crisis in healthcare AI is not going away, but it is forcing the industry to mature in useful ways. Companies are moving beyond the naive belief that more data automatically equals better models and are instead thinking carefully about the quality and economics of their data pipelines. This shift is healthy because it focuses attention on building sustainable businesses rather than just impressive demos.

The winners in the next phase of healthcare AI will be companies that treat curation as a core competency rather than an annoying operational detail. They will invest

expert networks, feedback loop infrastructure, and curation workflows with the intensity they invest in model development. They will recognize that in healthcare data curation is not something you outsource to the cheapest provider, it is a strategic advantage that compounds over time and creates defensible moats.

There are also broader implications for how healthcare AI integrates with clinical practice. If models are going to be continuously improved through feedback loops, we need new regulatory frameworks that allow for iterative updates while maintaining safety and efficacy standards. The current paradigm of treating AI models as static medical devices that get approved once and then rarely change is incompatible with the economic realities of maintaining high-performing healthcare AI. We need regulatory approaches that recognize the importance of ongoing curation and model improvement while still protecting patients.

For entrepreneurs building in this space, the message is to embrace the curation challenge rather than trying to engineer around it. Build your expert networks even before you have a product to annotate for. Create economic alignment with annotators through equity or revenue sharing. Invest in feedback loop infrastructure from day one rather than bolting it on later. And be honest with investors about curation costs and how they scale because the companies that succeed will be those that have thought carefully about these economics and built sustainable approaches.

The healthcare AI companies that win over the next decade will not necessarily be those with the most data or the most sophisticated models. They will be the companies that figured out how to access medical expertise at scale, built systems that learn continuously from clinical usage, and achieved sustainable unit economics through data curation. This is a harder problem than most entrepreneurs expected when they entered healthcare AI, but it is also a more defensible one. The barriers to entry created by expert networks and feedback loop infrastructure are real and lasting ways that pure algorithmic advantages are not.

The curation crisis is ultimately an opportunity for companies that understand the economics to build lasting competitive advantages. While others burn through capital trying to brute force the data quality problem, the sophisticated players are building

efficient curation machines that get better over time. That is where the real value in healthcare AI will be captured, not in having the biggest dataset but in having the most efficient path from messy data to clinical truth. For investors and entrepreneurs willing to think carefully about these economics, there are still enormous opportunities in healthcare AI. But those opportunities will go to the companies that respect the fundamental constraint that medical truth is expensive and building systems to generate it efficiently is the real innovation.

If you are interested in joining my generalist healthcare angel syndicate, reach out to treyrwales@gmail.com or send me a DM. We don't take a carry and defer annual for six months so investors can decide if they see value before joining officially. Accredited investors only.



4 Likes

← Previous

Next

Discussion about this post

Comments

Restacks



Write a comment...