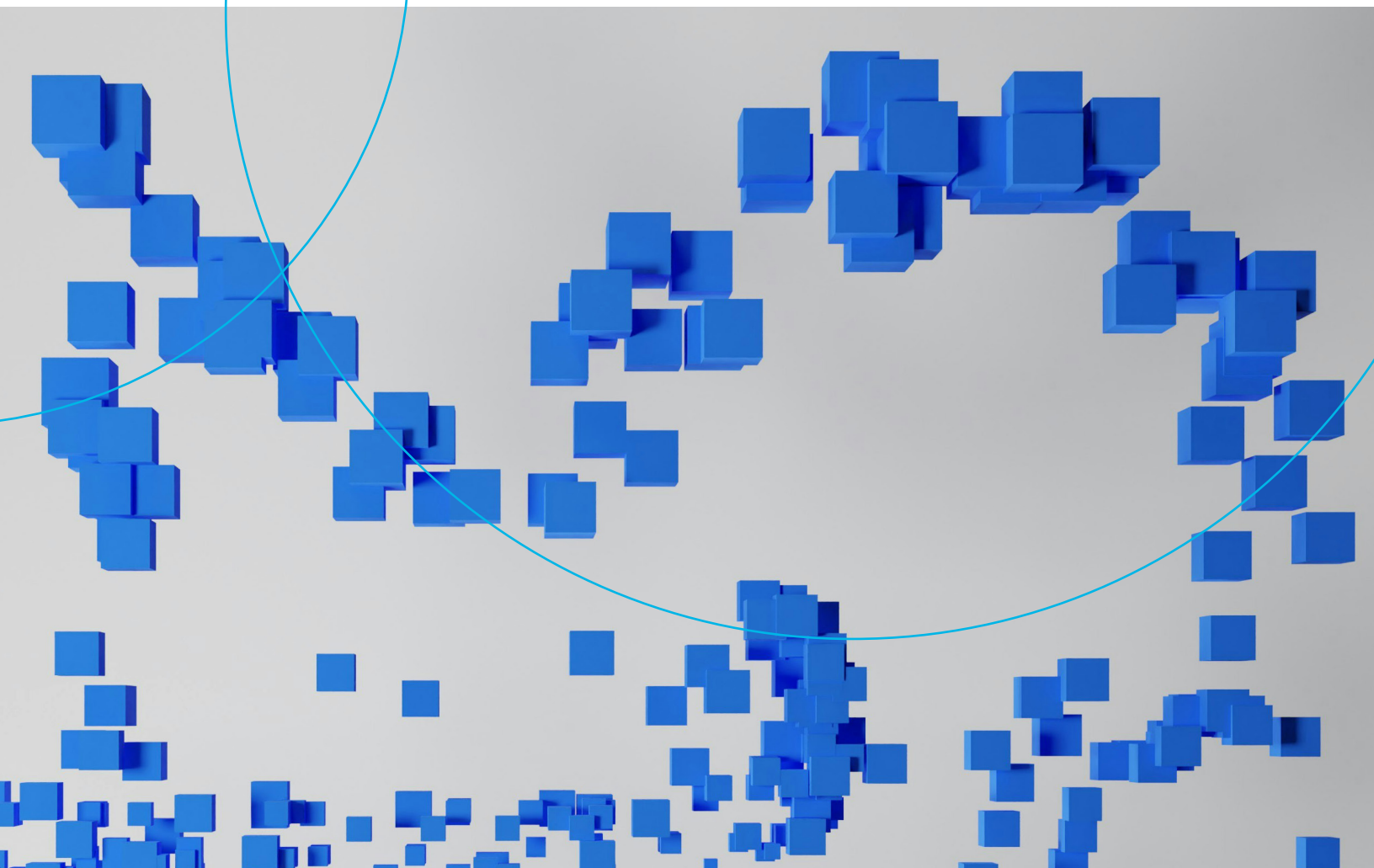
ThetaRho

White Paper
September 2025

How AI can solve healthcare's productivity crisis



The Why

Healthcare suffers from Baumol's Cost Disease meaning cost of labor has been rising without a concomitant increase in productivity. Although there is a lot of excitement in using AI for clinical use cases to increase repeatability and hence productivity in healthcare, we believe that focusing on non-diagnostic, clerical and administrative use cases can provide a better starting point with greater leverage and generalizability resulting in greater scaling potential for companies while addressing systemic cost problems in US healthcare.

Why? Because administrative costs in healthcare make up 25-30% of total cost.
Brookings Institution's Hamilton Project reports that the majority of this cost is in the actual healthcare service delivery and is borne by the providers. They tend to be non-profits or educational institutions but even the for-profit providers have small margins. They do spend a fair amount of money on IT because of various recent government mandates on sharing data and keeping records, but many have been disappointed with the returns on the investments they have made in, for example, EHR (Electronic Health Record) systems. The cost of implementing and maintaining these systems is prohibitive, but the users of these systems are extremely dissatisfied with them. Therefore, providers are a good place to start, especially if we can deliver quick ROI with limited CAPEX.

One key foundational problem is that current processes and systems in healthcare are cumbersome and complex because of regulation and fragmentation of market players. There are about 900 payors, 10,000 providers and 100,000 physician groups. There are many definitions of things they manage, especially critical things like reimbursement policies for clinical conditions, and this increases friction in the system. This compounded with multiple payors paying for a fee-for-service model produces perverse incentives, spiraling costs, and a lot of inefficiency.

Although a McKinsey report has identified about a quarter trillion dollars of potential savings in administrative expenses, not all problems are accessible to small companies. Some problems can be addressed within a single organization (payor or provider), some require collaboration across organizations (say between payors and providers), and some require systemic changes. We will focus on what can be done within a single organization, more specifically providers.

Even within administrative cost, non-billing and insurance related expenses are higher than billing and insurance related expenses according to The Brookings Institution. In addition, there are more than 250 companies in the revenue cycle management area of administrative expenses that focus on billing and insurance related expenses.

The non-billing and insurance related expenses are things like:
Querying and Updating EHR systems, quality measurement, customer service, scheduling, marketing, horizontal functions like HR, etc.

Specifically, there are many problems that individuals in these provider organizations face where AI-based technology can add value.

Clinical staff is experiencing burn-out at unprecedented rates, attrition is high and clerical/administrative overhead of using quaint EHR systems is a significant reason.

Current software systems supporting processes are not designed for clinical end-users but for billing purposes. This confuses users, slows down, and reduces quality of work, increases chances of mistakes and lowers capacity of the system. Not only that, but it is also expensive. Various estimates put the cost of replacing a physician between 500K and $1M.

On the other hand, patients see long wait times, poor experience as they navigate the healthcare system, particularly because of opacity of why and when things happen and how much they cost.

Order entry, coding for billing, scheduling, quality reporting, supply chain and customer service are other areas where sophisticated data-driven algorithms can help improve effectiveness of hospitals. While those are of eventual interest to us, we want to begin with a burning platform that will compel providers to act. In our conversations with providers, many seem to be looking for solutions for the burnout problem.

The What

Clinical staff burnout is a complex topic, but there are some common themes that are emerging based on our primary and secondary research.

Clinicians do a fair amount of clerical work prior to, during, and after encounters with patients. Some studies have shown that physicians spend twice as much time on clerical/admin tasks compared to direct patient interactions. In addition, during COVID, the amount of asynchronous interaction between patients and providers has gone up significantly. We now look at a few use cases, each of which represents a module with a clear end-user and value proposition for the user and the organization.

The "inbox" that was meant to provide a low-cost way to communicate has become packed with all sorts of messages that makes clinicians spend more than an hour each day on inbox processing where research has shown that a vast majority of these messages do not need physician input. Patients who generate a significant fraction of these inbox messages often expect a reply in a day, and organizations are not set up to handle the increase in inbox usage. AI, especially recent advances in NLP (Natural Language Processing) is uniquely positioned to solve this problem for providers and physicians while satisfying the governance, legal and policy requirements. There are common requests like prescription refills or test results from other doctors for the patient in the inbox for information only. These can be completely "automated" depending on the policies and procedures in each organization. Alternatively, imagine catching anomalous triggers for serious conditions by monitoring and interpreting messages in the inbox to improve primary care.

Computerized order entry of prescriptions, follow-up procedures, imaging, referrals, etc., is another area where data and AI can make a difference to make sure that there is consistency across patients and physicians. Not only to ensure that the drugs being prescribed do not interact with other drugs the patient is taking, but also that the insurance is likely to pay without prior authorization, etc.

Encoding various portions of the encounter for billing purposes creating say ICD10 or ICD11 codes is another area where AI can assist care providers by more accurately categorizing procedures and increasing the likelihood that they will get accepted by insurance providers on first submission.

Structured documentation generation is a key area where, for example, physicians must generate SOAP notes for each visit as well as visit summaries for patients. They either must spend time writing from memory, or they must record and use a speech to text technology to assist in this note generation. There are many players in this space including Microsoft, Oracle, and many other smaller players. Unlikely that we will get into this space at the outset.

Before arrival, physicians often want to scan the patient's history to understand their medical history and current patient-generated data to inform their line of questioning when they meet patients. Today, most EHR systems make it a laborious task to query patient history for pertinent info, and there is no chance of getting patient generated data. Sophisticated AI-based information retrieval solutions can provide a natural language interface as opposed to clunky form-based UI (User Interface). These represent possibilities for improving patient-physician interaction.

Sometimes, physicians are incented to fill in clinically tangential data about patients in the name of value-based care. This gives clinicians the feeling that they are data entry clerks, not people who are helping people live well.

There are other possibilities, but these give a flavor of the kinds of use cases we expect to provide in an experience centered around the clinical staff's training and mindset of treating patients, and the technology enables a deeper understanding of the patient's needs as opposed to being a hindrance and a time sink.

We first intend to focus on clinical staff (physicians, nurse practitioners and physician assistants) in specific areas such as primary care, internal medicine, emergency, ob/gyn, pediatrics and anesthesiology that are often the backbone of provider organizations and are close to the breaking point because of attrition and burnout.

However, the key here is that all these data and AI-driven solutions should not be looked at in isolation. Data silos are real everywhere and implementing AI on top of data within an enterprise and embedding them in enterprise applications is no simple task.

In healthcare, the silos for patient data are not just within a single organization but across organizations. Doing this integration upfront is key for success.

The How

There are a few considerations to be considered in the overall experience, solution architecture and deployment.

01. What will the end user experience be and where will it show up?
    a. We should re-imagine the experience for users, especially clinical users in providers.
        - Imagine, an intelligent inbox, that not only ensures that the messages get sent to the right folks, but it prioritizes messages and can generate draft automatic replies for clinical staff to use.
        - Imagine a natural language interface to EHR that allows the physicians to ask the question they want/need to ask of patient history and the machine finds the answers. Given the rapid advances in natural language processing, especially deep learning models like GPT, this problem is more tractable now than ever before.
        - It may not be possible to implement all the pieces needed here within the application's front-end/experience framework. In which case, we will create a layer on top, that provides the experience while keeping the system of record the same.

02. What EHR systems will we work with/integrate with and how?
    a. The top three EHR systems that account for about 80% market share are:
        - **Epic**
          Top market share and gaining among large, acute care hospitals.
          First four provider conversations we have had have been with Epic customers (Kaiser, Stanford, Sutter, UCSF)
          Perceived to be closed and difficult to work with
        - **Oracle/Cerner**
          Second in market share and losing share because of market uncertainty around acquisition by oracle.
          Likely to be busy with re-platforming onto Oracle stack. AI investments focused on note creation.
          More committed to openness
        - **Athenahealth**
          Popular among smaller physician practices, especially primary care. Has 160,000 users, cloud-based, open interfaces making it easier to get going.
    b. Given the increase in importance of FHIR, connecting directly is the way to go. We may have to get certified and have a "plugin" that needs to be installed in the host application.
03. What will be the deployment/distribution model?
    a. We will have to support an on-prem/virtual private cloud model.
04. How will we overcome the architectural and governance challenges healthcare IT organizations must overcome?
    a. Our data and AI platform will have to be "certified," e.g., HIPAA, SOC-2, etc.
05. What will be the customer journey from knowing they have the problem we can solve to solving it in production? Since we intend to be AI first, we will need to overcome the most common reasons why AI projects fail.?
    a. Overestimate or do not estimate feasibility and benefits. ROI for AI is difficult to calculate, and even if you did, it may change with time. Value depends on the data, workflow, and experience and all of them may evolve with time. Most enterprises do not know the "as-is" state of their data and workflows to estimate the impact of injecting AI on the user experience as well as business objectives. Once they have a view of the as-is state and a potential "end-state" the impact/ROI can be estimated. As it is, half the financial benefit for any transformation initiative is likely to be lost before implementation begins.
    b. Underestimate the effort and skill needed to succeed. Embedding AI into applications is much more complex than developing applications without AI. AI development has more moving parts, has more stages and is more iterative than traditional software. Any significant AI application has many models working in concert for business outcomes. AI solutions are also "brittle" in that changing anything changes everything. Having moving parts across SDLC (Software Development Life Cycle) can unknowingly introduce jitter and the experienced quality can materially differ from predicted quality. Cutting corners can introduce significant technical debt and lower quality outcomes.
    c. What is worse, when projects fail, there is little transparency in decision making around AI solutions and organizations are not able to learn from their mistakes. Organizations need to know why an AI solution was chosen, how it was designed and developed, on what grounds it was deployed, how it is monitored and updated, and the conditions under which it may be retired. Without transparency, governance is a non-starter.
    d. While there has been a lot of focus on modeling techniques, deep learning, etc. over the last 10 years, the data, the application logic utilizing the model and the eventual user experience are more integral to the success of AI solutions. The market has started moving in this direction with Data-centric AI.
06. However, most enterprises use applications and the experience these applications expose to customers and employees is how value from AI is delivered. What we need is Application-centric AI that can deliver a predictable experience and business value to enterprises. Here are some of the specific problems that need to be addressed:

a. The first step is a "discovery application" to estimate the potential of the AI solution based on the experience, workflow, and data available in the customer environment.
b. Data discovery is about finding out if the customer has enough quality data for the business problem at hand and what sort of effort is needed to create the data (Labeling, cleaning data, publicly available data, generated data, etc.). Ensuring that one has quality, representative data with provenance is a critical starting point for AI. Is the data used for model creation available at inference time and is there no "leakage."? These questions are domain and workflow dependent. Data quality is also not just a matter of misspellings, but if the data in the systems of record are reflective of "reality." The AI techniques we use must overcome the data challenges, as in the EHR case, data is often dirty, often copied and pasted, etc., and this can cause vanilla pre-trained models to not perform well.
c. Workflow/application discovery is about discovering how the set of AI models be used at various points in the application. Some of the key considerations are. Where will the prediction/decision be inserted? Is the data used to create the model available at inference time? What if data is missing? When will the application know that a particular inference call returned the "wrong" prediction? What is the cost of getting a "wrong prediction"? Can the application logic recover from a "wrong prediction"? What key business process metrics be affected and how does it depend on the "quality" of the AI model?
d. From an experience perspective, managing the change in the user expectations to build trust is key. This may be in the form of starting with a recommendation/suggestion or completely embedding the user experience with the application and "hiding" the AI from the user. Providing ways to gradually expand the set of users who are exposed to AI so that the experience can be predictable before adoption at scale.

07. Development is much more iterative than traditional solutions. Because value is dependent on the data, it is important to stand up what is the eventual solution end to end quickly and iterate. Quality improves with fast iterations. But organizations may not have the patience because the improvements are not likely to be predictable. Shortening the time to do an end-to-end run will help increase chances that customers go through to a place with enough quality to adopt at scale. This means that the entire SDLC for AI must be repeated often for each application. This requires CI/CD on steroids across experimentation, development, testing, deployment, and management. Testing continues well into production. It is not as simple as "detecting data drift" and retraining. Because user experience may change unexpectedly with new models. AI systems are brittle in that changing anything changes everything, and there are lots of things to change to "improve quality." Therefore, all the related pieces must be put together to deploy as a unit.
08. Finally, computer workloads for AI are much more diverse and difficult to characterize than traditional workloads, scale with data and other factors, so getting predictable runtimes is harder. This makes knowing how much infrastructure to use difficult and expensive. On the other side, workloads (both data pipelines, model creation and serving) are often parallelizable, but most folks do not take advantage of it, because they do not have the technical expertise to put these systems together. Support is much more difficult for products with AI in it, and platforms must have inbuilt observability and debuggability for support.
09. The notion of systems of intelligence captures what is needed in theory, but doing it in practice is tough because the sheer number of moving parts and the technical complexity trips up most enterprises.
10. We want to solve this complexity so that healthcare provider organizations can benefit from AI. If we succeed, we may be able to tackle other verticals as well. Healthcare is where we are starting because:
    a. there is a real burning problem where AI can make a difference,
    b. the macro trends and government mandates in the industry are making it more open and The expectations from the younger generation of physicians are vastly different. They expect consumer grade experience at work and the current technology systems they deal with fall noticeably short.
    c. We want to ride these macro trends.