

Beyond Primary Care: Unlocking ROI for AI Through Coding Support and Specialty Expansion



In 2024, the healthcare industry welcomed a bounty of AI-powered tools, seeking solutions to its persistent and growing challenges. Industry leaders and stakeholders flocked to AI solutions driven by a central thesis: empowering people with AI could significantly enhance the quality of care, reduce healthcare costs, and ultimately deliver a better experience for patients, families, clinicians, and all healthcare workers.

This adoption comes at a critical time: healthcare systems have been forced to operate with widespread staffing shortages, with an estimated deficit of 125,000 clinicians by 2025. Simultaneously, the demand for care continues to rise, with approximately 11,000 seniors aging into Medicare every day.

On top of that, our current healthcare system is plagued by inefficiencies. Physicians spend only about a quarter of their workday on direct patient care. The rest of their time is consumed by an ever-growing mountain of paperwork, data entry, and administrative responsibilities that detract from their primary mission: treating patients.

This problem significantly contributes to burnout and attrition among healthcare professionals, creating a vicious cycle that has long plagued the industry. As burned-out clinicians leave the field, the workload for those who remain intensifies, perpetuating a self-reinforcing crisis in healthcare staffing and quality of care.

The industry first embraced off-the-shelf AI models for healthcare, but quickly realized that models designed for generic use cases were ill-equipped to handle healthcare's specific demands. Clinical leaders quickly learned that AI models needed to be tuned specifically for healthcare. Secondly, AI models would need to be carefully designed for each specialty and subspecialty.

Healthcare data inputs are messy; excess noise, multiple speakers, accents, and varying pronunciations of medications make it difficult for general models to know what's going on.

Furthermore, documentation also needs to accurately reflect clinical decision-making, which clinicians may never directly verbalize. The model needs to have the capability to read between the lines and fill in the blanks.

General models are designed to be excellent at tasks like summarization, but struggle when it comes to documentation in the medical context and lack the specialized knowledge needed to perform healthcare tasks reliably.

For healthcare AI models to be effective, they need more than just textbook knowledge. They must be deeply integrated into the medical record context such as previous health issues, be designed for each specialty and subspecialty, and have a thorough understanding of the internal operations of the health system (coding, prior authorization).

Otherwise, it's like placing a well-read college student in a hospital and expecting them to be immediately useful. It won't be able to solve healthcare tasks reliably and will make mistakes that are unacceptable in a medical context.

While we might hope that healthcare is solely about patient care, the reality is that physicians need to know more than just medicine to operate effectively in the healthcare system. They must understand coding, prior authorization, and keep track of clinical trials. This is why 73% of a physician's time is spent on nonclinical tasks. To alleviate this burden with technology, AI models must be trained to comprehend these additional responsibilities as well.

REDUCING BURNOUT AMONG SPECIALISTS

Clinician burnout is particularly prevalent among specialists, who face prolonged work hours, high patient loads, and the need to stay updated with rapidly evolving medical knowledge. The pressures to achieve precise diagnoses and effective treatments, often for patients with complex or rare conditions, further accelerates burnout.

Specialists must also navigate complex coding and billing systems, manage extensive documentation requirements for prior-authorization, and coordinate care with multiple providers. These administrative tasks not only consume valuable time but also contribute to burnout, as they divert specialists' attention from clinical work and reduce the time available for direct patient care.

At the same time, specialists are an important factor in the financial health of health systems due to the high value and complexity of the services they provide. Specialist expertise in diagnosing and treating complex conditions often leads to higher reimbursement rates from insurers, particularly for procedures and treatments that require advanced skills and technology. Specialists also drive referrals and attract patients to hospitals and clinics, increasing overall patient volumes and revenue through associated services like imaging, lab tests, and surgeries. Furthermore, specialized care also enhances a health system's reputation, drawing in patients from a wider geographic area.

BUILDING AI MODELS FOR EACH SPECIALTY AND SUBSPECIALTY

When evaluating AI solutions, it is important to understand how models are configured and customized. If done correctly, a team of engineers will tune an AI model to be fully contextually aware of the medicine, workflows, and nuances of each specialty and subspecialty. If done incorrectly, a solution will leverage a single generic model and try to broadly apply it across all clinical specialties. As you can imagine, the resulting documentation often requires substantial editing by the clinician. In these instances, what is generated also often ends up being incompatible with the coding and billing idiosyncrasies of each unique service line.

At the highest level, the overall structure of documentation looks different between different specialties and subspecialties. For example, documentation for Orthopedics might have a section dedicated to review past imaging. Cardiologists will want an entire separate section dedicated to review past electrical physiological studies. Psychiatrists will need one set of documentation for medication management, and a completely separate set of documentation for psychotherapy.

But even within the same sections of documentation, what matters to a primary care physician is different from what matters to a specialist. Therefore, how content is captured and organized must be different. The example below demonstrates the differences between an HPI section generated by a general vs. specialty-tuned model.

Beyond Primary Care: Unlocking ROI for AI
Through Coding Support and Specialty Expansion



Subspecialty-specific tuning (Orthopedic hand surgery example)

General Visit HPI - Not Specialty Tuned

HPI: Patient reports waking up in the morning with tingling and numbness in the left hand, as well as difficulty picking things up. Patient also reports stiffness in the index and ring fingers, as well as the thumb. Symptoms have been present for 3-4 months and have not improved with Advil.



Insufficient detail on:

- Grouping specific hand-related symptoms by underlying problem (numbness & tingling vs. stiffness).
- Specific details on fingers affected.
- Exacerbating and alleviating factors.

Ambience-tuned HPI from the same visit

HPI: Patient reports a three to four month history of numbness and tingling in their left hand, specifically in the pointer and ring fingers. The numbness and tingling are not constant, but rather come and go, with symptoms being worse in the morning. The patient also experiences these symptoms when driving or holding a phone. The patient has not had a nerve conduction study.

In addition to the numbness and tingling, patient also experiences stiffness in the same fingers, as well as occasionally in the thumb. The patient describes the stiffness as a lack of fluidity in the movement of these fingers. The index finger has been observed to lock down in a certain position and then pop straight. This has been occurring for the same duration as the numbness and tingling.

Patient has been attempting to manage symptoms with Advil, but reports no improvement. The symptoms sometimes wake the patient from sleep, approximately twice a week. They have not tried wearing a wrist brace at night.

In the example above, the general AI model was not able to generate usable documentation for Orthopedics. Key information regarding the patient's symptoms was left out, and the information that was captured was not organized correctly.

Another example below shows the difference between a general AI model vs. a model designed specifically for Ophthalmology. The general AI model did not understand the importance of capturing the patient's full family history of retinitis pigmentosa, and did not include the patient's past diagnostic results. However, the AI model specifically designed for Ophthalmology provided comprehensive documentation of the patient encounter.

Specialty Specific Tuning (Ophthalmology - Retina HPI Example)

General Visit HPI

Patient was diagnosed with retinitis pigmentosa at age 30, and now at age 70, reports difficulty with her vision and increasing challenges in her work as a visual artist. She has a family history of retinitis pigmentosa, with an older sister also affected. Patient underwent cataract surgery in both eyes at age 36, which improved her vision. She has never worked with a low-vision specialist.



Insufficient detail on:

- Length of vision decline in vision and details on impact on patient's career.
- Family history of RP in older sister, and absence of RP in brother and parents.
- Cataract surgeries at ages 30 and 35 (not 36, and not completed at the same time)
- Lack of experience with low-vision specialists or adaptive vision aids aids.
- Past diagnostic result findings (genetic testing results)

Ambience-Specialty tuned HPI

70-year-old female with a past medical history significant for RP diagnosed at age 30 and cataracts diagnosed at the same time, presents for evaluation of her vision. She first noticed vision problems during pregnancy at age 30, which led to the diagnosis of cataracts and a "mild case" of RP. Over the past 40 years, her vision has progressively declined, with a significant decline noted about 1.5 years ago with her peripheral vision, severely impacting her lifelong career as a visual artist. She recently applied for disability due to her vision, but was denied because she was self-employed.

She has a family history of vision problems. Her sister, who is 11 years older, also has RP. Her brother does not have RP. Her parents, now deceased, did not have known retinal degeneration. Patient underwent cataract surgery at ages 30 and 35, which significantly improved her vision at the time. She has not sought help from a low-vision specialist or utilized adaptive aids for her vision.

Diagnostic Results:

- Genetic testing (April 17th, 2023) - Results mailed to Dr. [Name], indicating a change associated with the gene SNRNP200, inherited in an autosomal dominant manner.

22

Beyond Primary Care: Unlocking ROI for AI
Through Coding Support and Specialty Expansion

Both examples demonstrate the importance of designing AI models that understand the unique medicine, workflows, and documentation requirements of each specialty and subspecialty. If a specialist is looking at documentation and it looks like primary care documentation, they will either waste time editing the note or stop using the technology entirely. In this instance, the health system will have a difficult time reaching widespread adoption.

THE CODING CONNECTION

CLINICAL USE CASE: MEDICAL CLAIMS

A major part of clinicians' administrative burden is keeping track of and adhering to coding rules, which change from year to year. Medical codes are an essential part of clinical documentation and have a direct impact on payor reimbursement, prior authorization approval, clinical trial matching, and proper continuity of care. They help health systems understand how to appropriately allocate funds to ensure sufficient resources for patient care.

However, clinicians are not coders. They often make mistakes by selecting incorrect codes or failing to include the necessary details in their notes for payors to approve their documentation. A study by AAPC Services found that coding errors result in the denial of 24 percent of claims which exhausts clinical and administrative resources.

Medical coders review claims to ensure that all codes reflect the severity of the clinical issue. However, coders are not clinicians and their ability to catch every improper or missing code is limited. Undercoded claims lead to an annual loss of nearly \$15,000 per clinician.

When codes are incorrect or missing in a patient's chart, it starts one of two processes. It either results in a rejection of the chart, which must require the clinicians and staff to go through an appeal process or, if a medical coder catches it, it instigates a Clinical Documentation Improvement (CDI) query, which sends the chart back to the clinician to correct their mistake.

Consider, for example, a diabetes patient going to see their physician for a foot ulcer. The physician and patient discuss the "wound" and an AI scribe would document accordingly with the diagnosis code of E11.621: Type 2 diabetes mellitus with foot ulcer.

However, if the model didn't change the word "wound" to "ulcer" in the clinical notes, the patient's chart would be subject to a CDI query or rejection from a payor. Payors would also look for the exact location of the ulcer which requires an additional code: E11.621 (Non-pressure chronic ulcer of left heel and midfoot).

In order to achieve clinical notes that contain both necessary codes, the model needs to be trained with a comprehensive understanding of medical terminology, context, and coding requirements. This ensures that the documentation is precise and complete, minimizing the risk of rejections and CDI queries.

CLINICAL USE CASE: PRIOR AUTHORIZATION

Patients who need advanced, complex procedures often require approval from their insurance companies, which is referred to as prior authorization. Payors look for two things: clear diagnostic evidence of the condition and a record of cheaper, easier solutions not working.

Beyond Primary Care: Unlocking ROI for AI
Through Coding Support and Specialty Expansion



When prior authorizations are denied due to improper or incomplete coding, physicians then have to spend their time on the phone with the payor to explain why the procedure they recommend is medically necessary.

Conversely, if everything is done correctly from the start, the physician spends less time disputing with payors later. This ensures that patients receive the care they need promptly.

If a model understands what payors require in documentation to prove medical necessity, then physicians can avoid spending their time disputing claims with payors.

NEXT-GENERATION HEALTHCARE AI MODEL USE CASES

While coding-aware models address the administrative burdens during and after patient care, what can be done about the challenges faced beforehand?

Pre-charting is when clinicians prepare documentation before a scheduled appointment or encounter. When a patient has complicated medical history, this can be where a clinician spends the majority of their time.

Pre-charting is more challenging for AI models than post-visit documentation. These models must navigate through extensive historical patient data and correctly identify and amalgamate crucial clinical information that is relevant to the immediate concern. Unless clinicians meticulously review all of the data themselves, they may lack the ability to safely audit the AI's decisions.

As a result, the quality standard for pre-charting models is exceptionally high, far exceeding that required for documentation and coding. In those areas, physicians can easily verify the information due to their expertise and access to detailed data. However, pre-charting demands a much higher accuracy because it involves synthesizing complex information that the physician may not have readily available to cross-check.

For example, consider a cancer patient with suspicious lung nodules. An oncologist needs two critical elements: the T&M staging and the ECOG status. The T&M staging is essential for understanding the advancement of cancer. This staging is crucial for making informed decisions regarding treatment and estimating the patient's prognosis. Without it, the oncologist cannot accurately gauge the severity of the cancer.

The ECOG status measures how effectively a patient can manage their daily life. This status is vital in helping the oncologist determine the aggressiveness of the treatment approach. If a patient is already struggling with daily activities, more conservative treatment options are necessary. Conversely, if the patient is thriving at home, a more aggressive treatment approach can be considered to tackle the cancer more effectively.

However, obtaining the T&M staging and ECOG status is challenging because the model needs to perform two complex tasks. First, it must retrieve multiple discrete pieces of information from the patient's chart, which may be buried within extensive medical records. Then the model must synthesize this information using specific formulas that are nuanced and vary depending on the type of cancer. For instance, the staging process for lung cancer differs significantly from that of cervical cancer or other cancer types.

To address pre-charting challenges, next-generation clinical AI co-pilots must master the ability to identify important data and synthesize it accurately and reliably. Everyday we get closer to manifesting this potential.

In fact, healthcare AI models are becoming increasingly multimodal, which enhances their performance across various clinical use cases. More context equals more capabilities.

Beyond Primary Care: Unlocking ROI for AI
Through Coding Support and Specialty Expansion



For example, two teams at a major tech company were working on large language models (LLMs) for different tasks – one model was for writing software and the other model was for answering general questions like a chatbot. They decided to see what happened when they combined the models. Despite using the same data and compute resources, the new model outperformed the previous model at both tasks. With more knowledge and context, the model showed more common sense and reasoning abilities. Healthcare models show a similar trend: improvement in documentation leads to better performance in pre-charting, coding, and vice versa. Multimodal models could potentially operate as AI co-workers on every care team. Imagine an AI that could automatically follow up with patients between visits, ensure they adhered to treatment plans, picked up medications, and got necessary tests, all while monitoring symptom progression.

GOVERNING AI: DEFINING QUALITY

In order to evaluate healthcare AI, we must define quality. Defining quality in healthcare AI models is more challenging than it may initially appear. Typically, when a model fails in a clinical setting, it usually boils down to one of two root causes. The first is that the model either lacks the requisite knowledge or reasoning capability to solve a particular problem. This is relatively easy to define.

The second, and more difficult issue, is determining whether these models are aligned to behave in a way that is useful and safe.

Understanding precisely what an oncologist needs and what is important to their decision-making process is a much more complex question than simply verifying if the model knows a specific clinical fact.

This issue isn't unique to healthcare models. The concept of releasing a model specification (model spec) is straightforward: developers and AI specialists must document the exact framework for how models should behave. Clarity on model operations is essential, and it is crucial to foster an open dialogue among stakeholders to continually refine this framework.

Defining quality involves answering questions like:

- What should the model's objectives be?
- What rules should it follow?
- What are good default behaviors?

Healthcare-specific model specs must consider the various ways clinical AI capabilities will be used in practice, the nuances of quality in different care settings and specialties, the scope of practice for different user types, and how to ensure models do not exhibit harmful biases before deployment.

GOVERNING AI: HEALTHCARE NEEDS ROBUST EVALUATIONS

Once a definition of quality is agreed upon, it must be measured. While a model spec helps define quality, evaluations (evals) are crucial for assessing whether a model instance meets the necessary quality standards before deployment.

The eval concept is elegant because it addresses a significant challenge in AI governance.

Beyond Primary Care: Unlocking ROI for AI
Through Coding Support and Specialty Expansion



Governing AI is difficult due to the constant evolution of inputs like training data, architecture, and hyperparameters. It remains a research problem to link changes in model behavior to changes in these inputs. Many AI governance initiatives focused on tracking inputs have failed because it's hard to draw meaningful conclusions about a model based on its training process.

Evals simplify governance by focusing on what truly matters: the model's behavior and its alignment with the model spec. This approach ensures that models meet the desired quality standards regardless of their training specifics.

THE PROFOUND IMPACT OF CLINICAL AI WHEN DONE RIGHT

It's essential to recognize the profound impact these technologies can have when implemented correctly. For organizations, AI can streamline operations, reduce costs, and improve resource allocation. However, in order to achieve these results, AI models must be built to understand the medicine, clinician workflows, and documentation requirements for each specialty and subspecialty. This will allow clinicians to benefit from decreased documentation burden, allowing them to focus more on patient care and less on administrative work. Only then can patients experience more timely and accurate diagnoses, personalized treatment plans, and overall enhanced care quality. When it comes to realizing the potential of AI in healthcare, we are only at the beginning.

Beyond Primary Care: Unlocking ROI for AI
Through Coding Support and Specialty Expansion



About



Ambience Healthcare's AI technology for scribing, coding, and patient summaries has been deployed at leading healthcare organizations including UCSF, St. Luke's Health System, John Muir Health, and Memorial Hermann Health System. Ambience is the only AI scribing and coding solution designed to support 100+ specialties and subspecialties and is directly integrated with leading EHRs.

By partnering with Ambience, healthcare systems reduce documentation time by an average of 80%, improve coding integrity, and achieve at least a 5X return on investment with more accurate E&M coding.

[LEARN MORE](#)



Beyond Primary Care: Unlocking ROI for AI
Through Coding Support and Specialty Expansion

