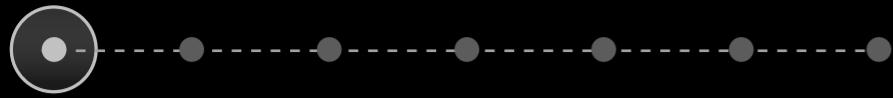


Innovation at the Speed of Trust

Contents

01	Introduction	03
02	Specialty Selection: Data-Driven Prioritization	04
03	Development: The Clinician-in-the-Loop Model	08
04	Testing: The Multi-Layered Validation Stack	10
05	Intentional Rollout: The "Early Wave" Strategy	16
06	Scaling to GA: Silent Excellence	18
07	Conclusion: The Future of Iterative Intelligence	22



Introduction

In December 2025, Abridge rolled out several new specialty-specific model updates, advancing from adaptable note-taking to deeply distinctive, clinician-aligned documentation tailored to highly specialized workflows.

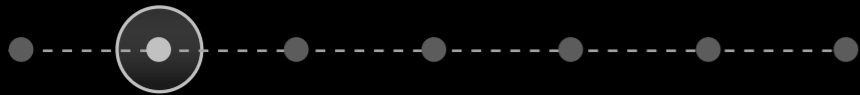
These updates are developed and rigorously evaluated to ensure measurable gains in note quality before they are released into the hands of clinicians, which is all part of Abridge's continuous improvement process for clinical AI.

When these model updates are ready, they are shipped "quietly," which is to say without build, configuration, or workflow disruption. Health systems are not asked to implement new versions. Clinicians are not prompted to update apps. The process is seamless, enhancing care delivery without interrupting it. Clinicians simply experience better notes, automatically.



Specialty Selection: Data-Driven Prioritization

We don't assume which specialties need attention;
we listen to clinical demand.

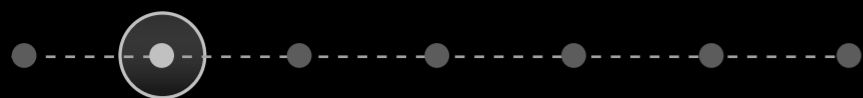


Even small details, like in what circumstances the phrases 'history of' or 'severe' is appropriately used in a note, can have real downstream impact on billing, liability, and the time clinicians spend finalizing a note.

This is why every specialty model update starts with a thorough definition phase, where we analyze many real life conversations and encounters to understand the baseline behavior of our models, spend hours speaking to clinicians and their hoped-for improvements, and establish consensus on the meaningful changes to implement.

This process of deep analysis of real notes, guided closely by feedback from our clinical partners, is what powers ongoing improvements across all our models so that Abridge today is always better than Abridge yesterday.

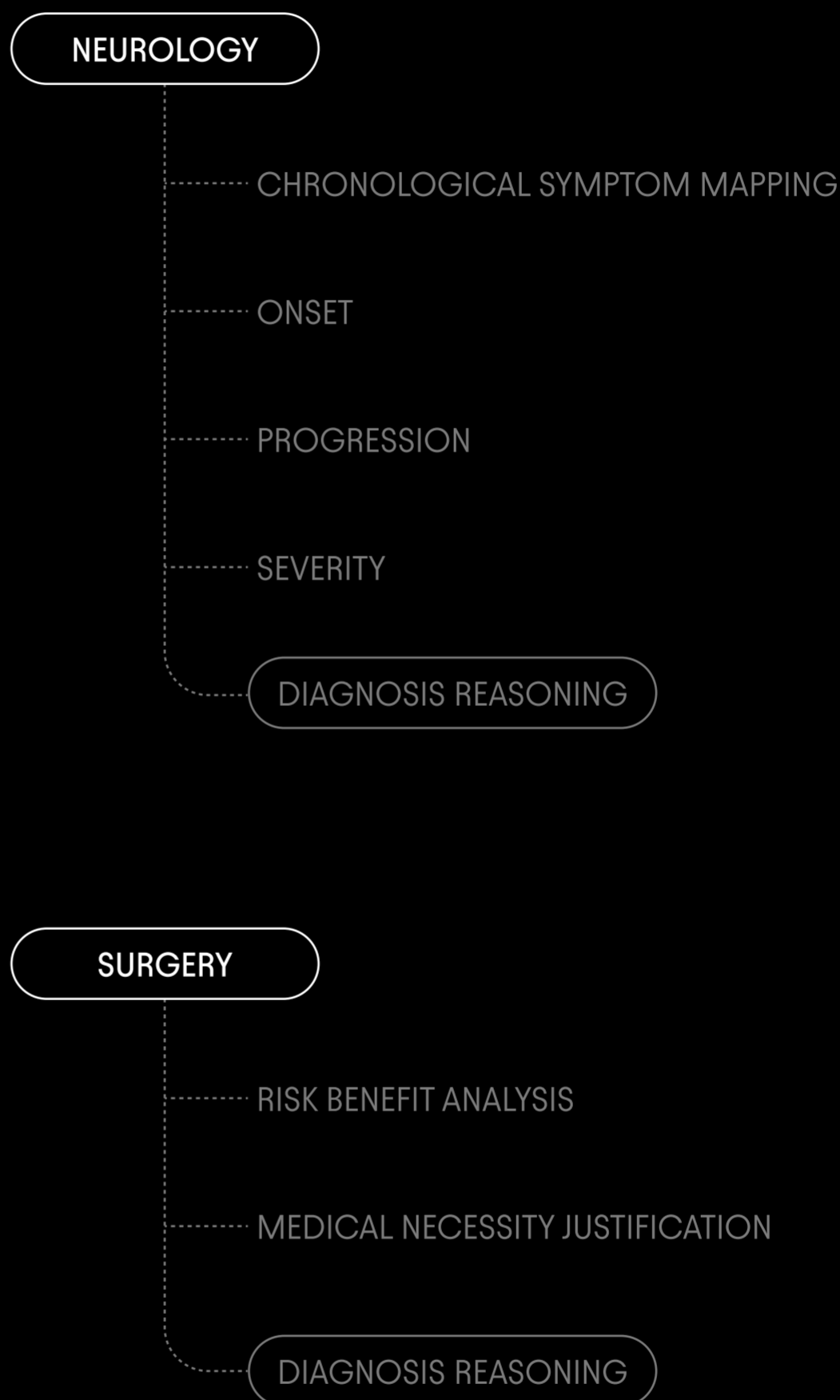
Katherine Choi
Senior Product Manager
Abridge

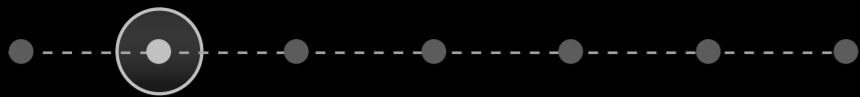


At Abridge, our documentation models support clinicians across specialties, care settings, and systems at scale. It's that scale that gives us the opportunity to go a million miles deep in specific specialties to make them even better.

For example, neurology documentation demands precise chronological symptom mapping, capturing onset, progression, and severity over time to inform diagnostic reasoning and treatment decisions. While surgical documentation requires explicit documentation of risk-benefit discussions and clear medical necessity justification.

Each specialty inherently requires specificity, and high-quality documentation must reflect the distinct logic of its respective clinical workflow.



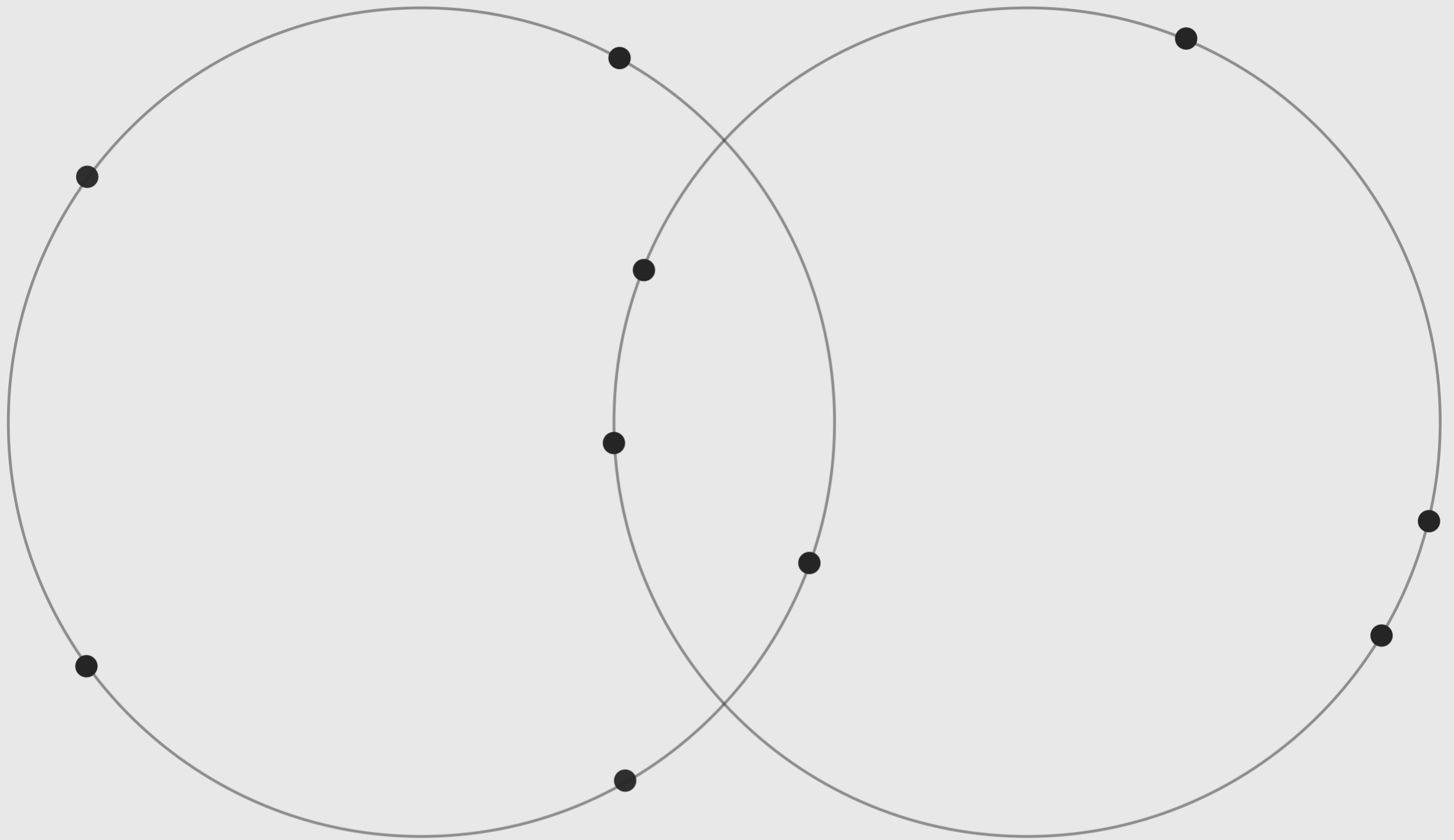


Specialty models are tuned to structure and prioritize those findings correctly, often in tighter, more structured note formats with terminology specific to each specialty.

Furthermore, these formats also need to be tailored to user preferences, which vary widely within specialties: from comprehensive, to concise, to bulleted across patient narratives, all with clinical judgement represented throughout. And the best approach to aligning note formats with user preferences is simple: let clinicians personalize their notes.

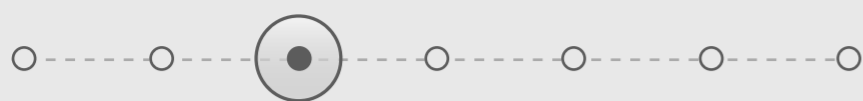
Giving clinicians agency over their note structure, refining notes with simple language prompts to adjust for tone or add specificity, is one way Abridge powers efficiency. Next, Abridge will introduce contextual prompts at the point of conversation, surfacing relevant insights in real time during clinical conversations without interrupting workflows.

All of this is based on real world insights. By grounding prioritization in clinical demand and feedback, we can identify where a one-size-fits-all model may not capture the realities of a given specialty. When we see friction, that's our signal to investigate and ensure our specialty updates solve challenges clinicians actually experience in their daily practice.



Development: The Clinician-in-the-Loop Model

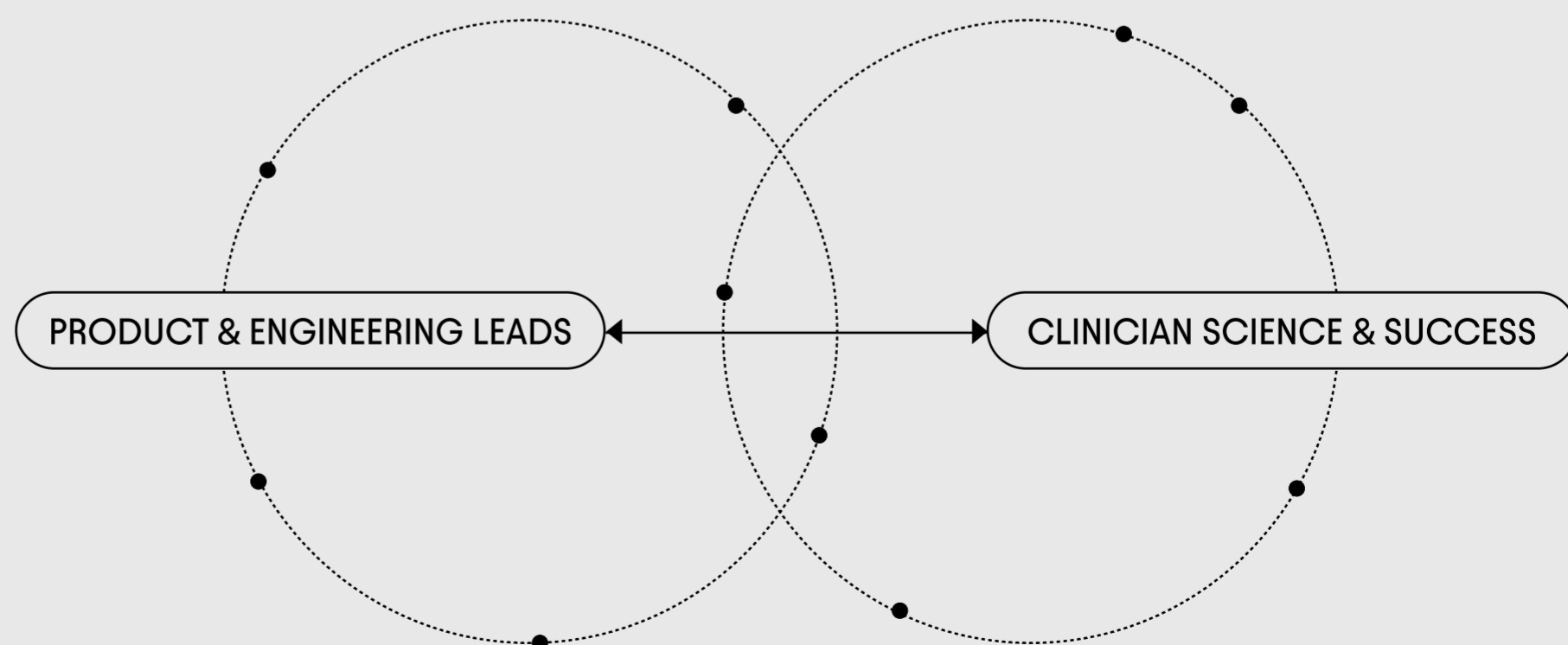
Building models that speak the language of each specialty requires deep collaboration.

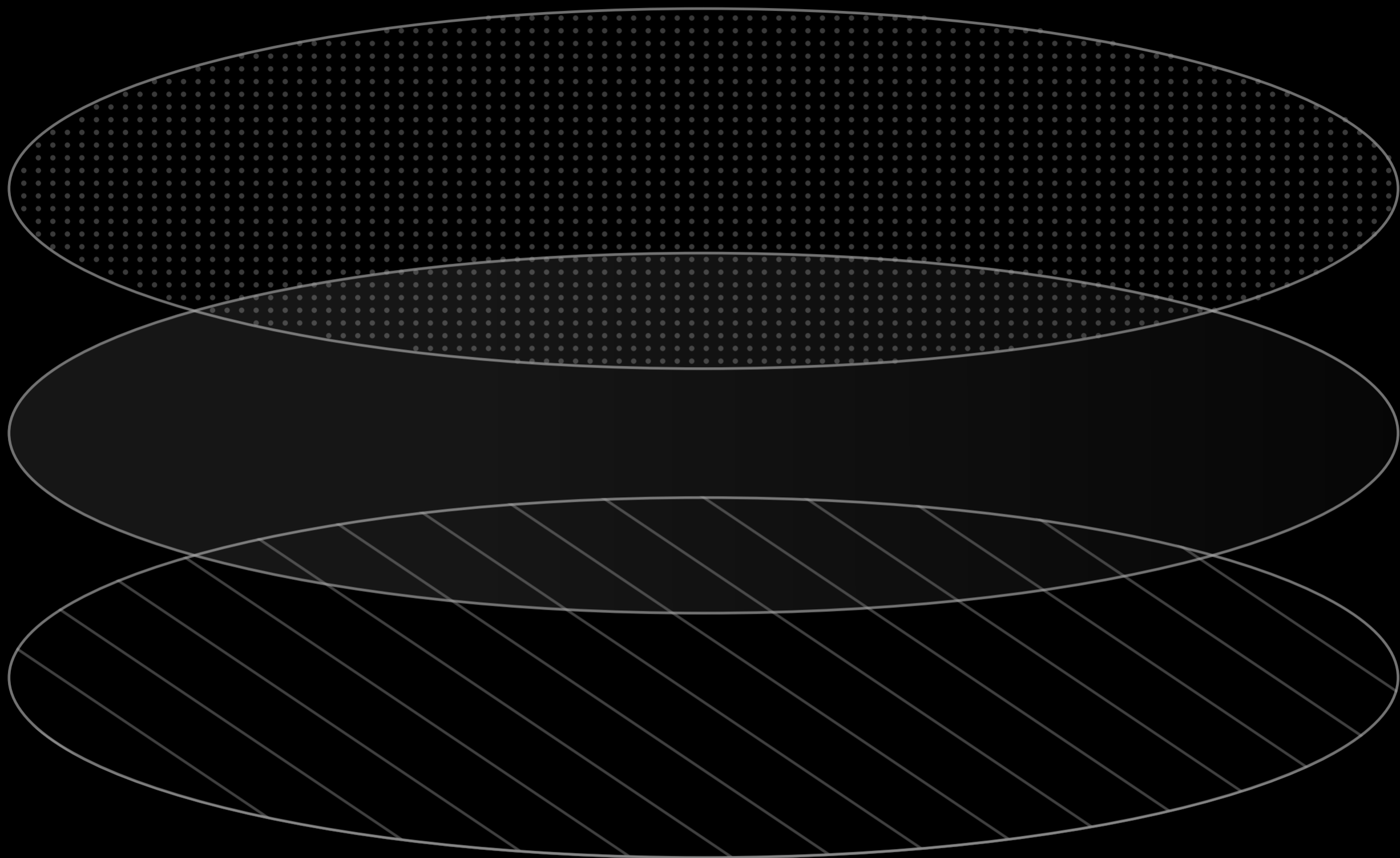


Once a specialty is prioritized, development begins. Our “clinician-in-the-loop” process pairs product and engineering leads with Clinician Science and Clinical Success Directors in cross-functional collaborations we call “NoteGen” teams.

Together, we refine the underlying “recipes” that shape how different sections of notes are generated. The goal isn’t only accuracy; it’s alignment. We build these models to mirror not just how specialists document care but also how they think about care.

By developing alongside clinicians, we ensure specialty updates not only reflect real workflows but also feel familiar. It’s a collaborative process we continue to strengthen, with “recipes” we continue to refine.





Testing: The Multi-Layered Validation Stack

Before a single line of code hits a healthcare system, it passes through a rigorous "gauntlet" of evaluations.



When model updates are in development, we pressure-test quality, accuracy, and clinical flow through a rigorous three-layer validation process that includes hands-on clinician review, third-party audits, and automated model evaluation.



01

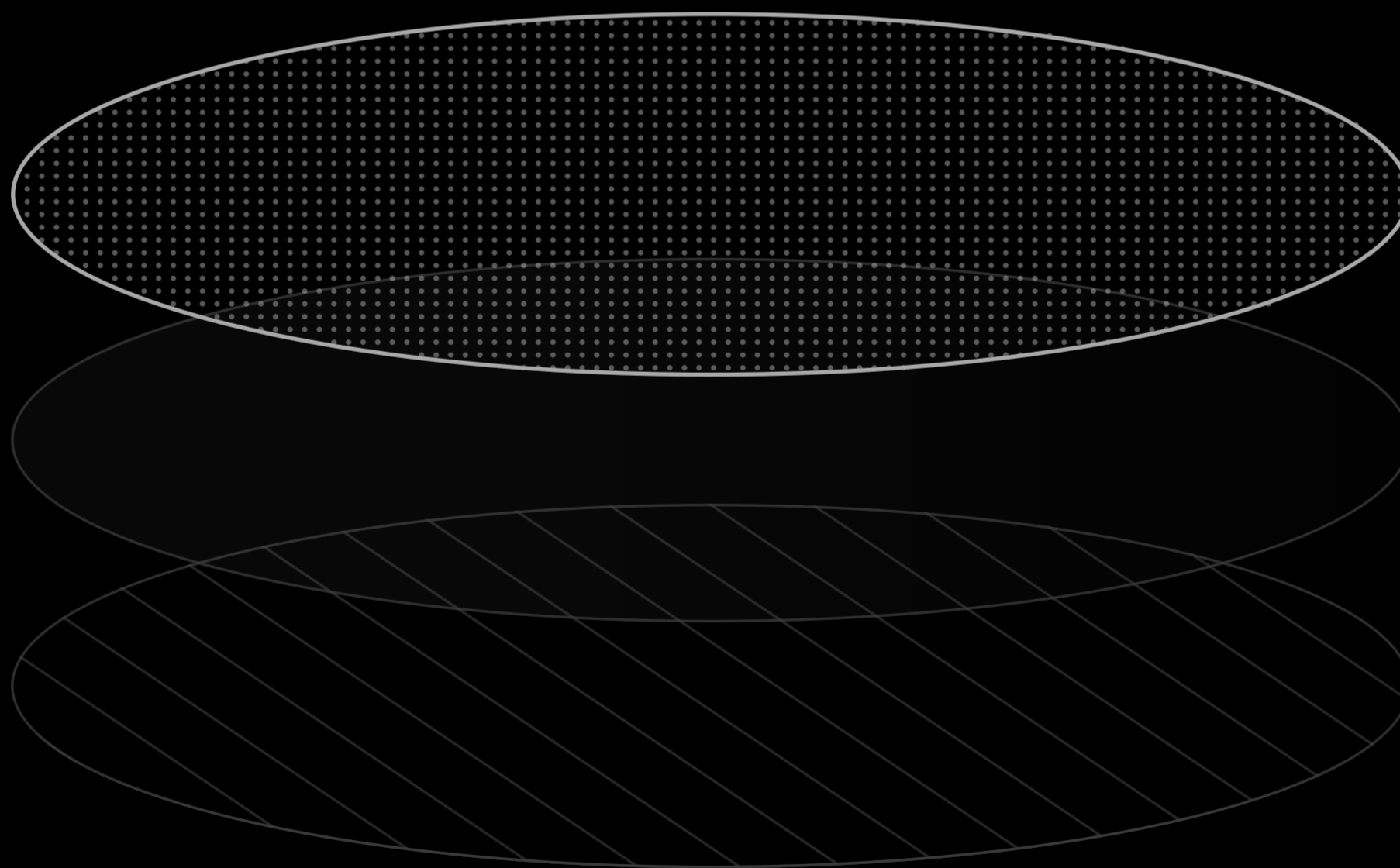
To start, we apply an automated LLM evaluation, what we call the "Judge" layer. These "judges" score the new model notes on a number of different variables, including:

NON INFERIORITY

Strict safeguards that ensure new specialty models match or exceed the generic baseline performance.

MISATTRIBUTION RATES

Monitoring against "hallucinations" or misattributed patient data



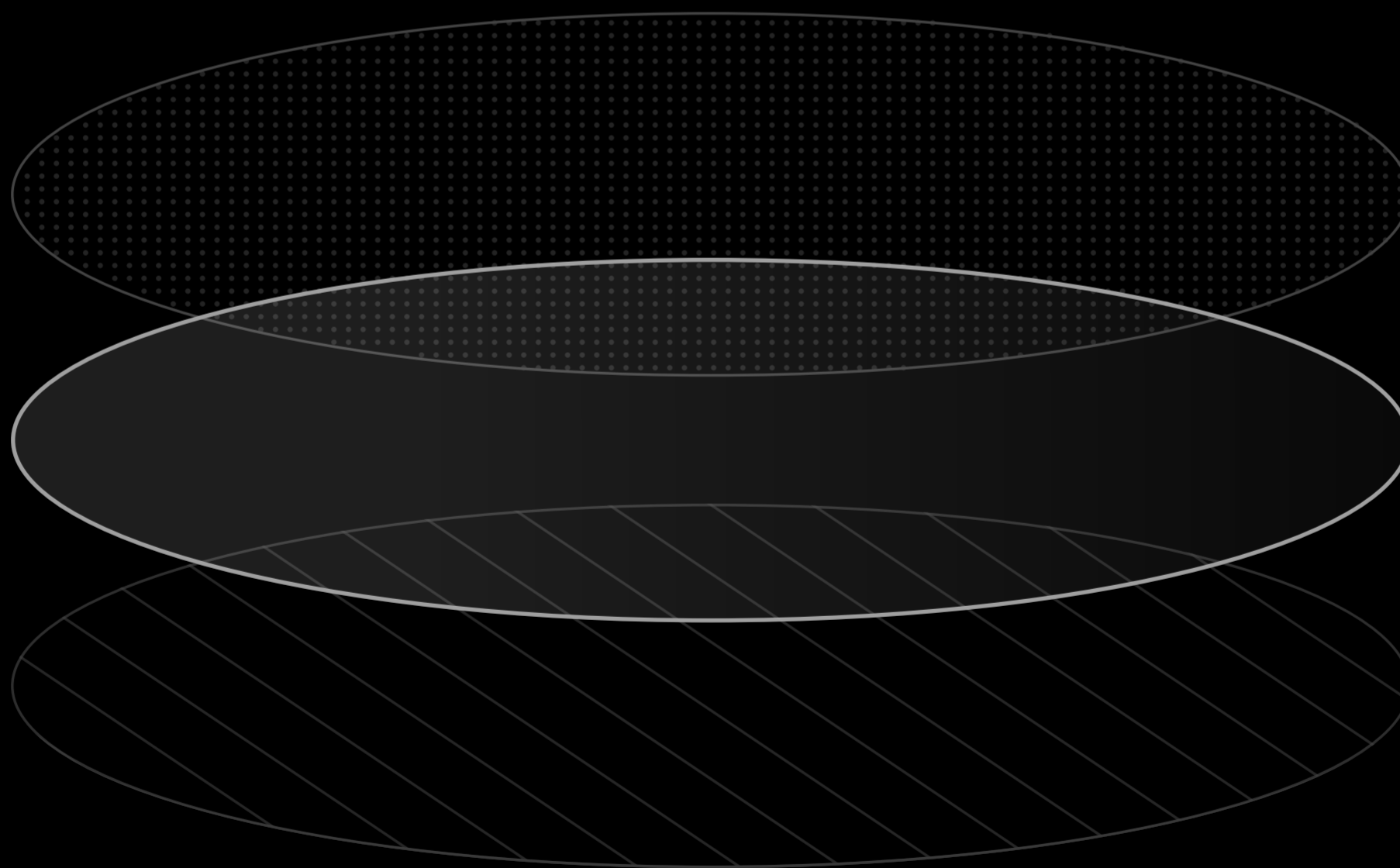


02

Then our Clinical Success Directors evaluate real-world examples against a predefined, data-driven baseline to determine if the update meaningfully improves performance.

CLINICIAN-ADJUDICATED VALIDATION

Licensed clinicians evaluate real-world notes against predefined, data-driven baselines to confirm measurable improvements in quality and accuracy.



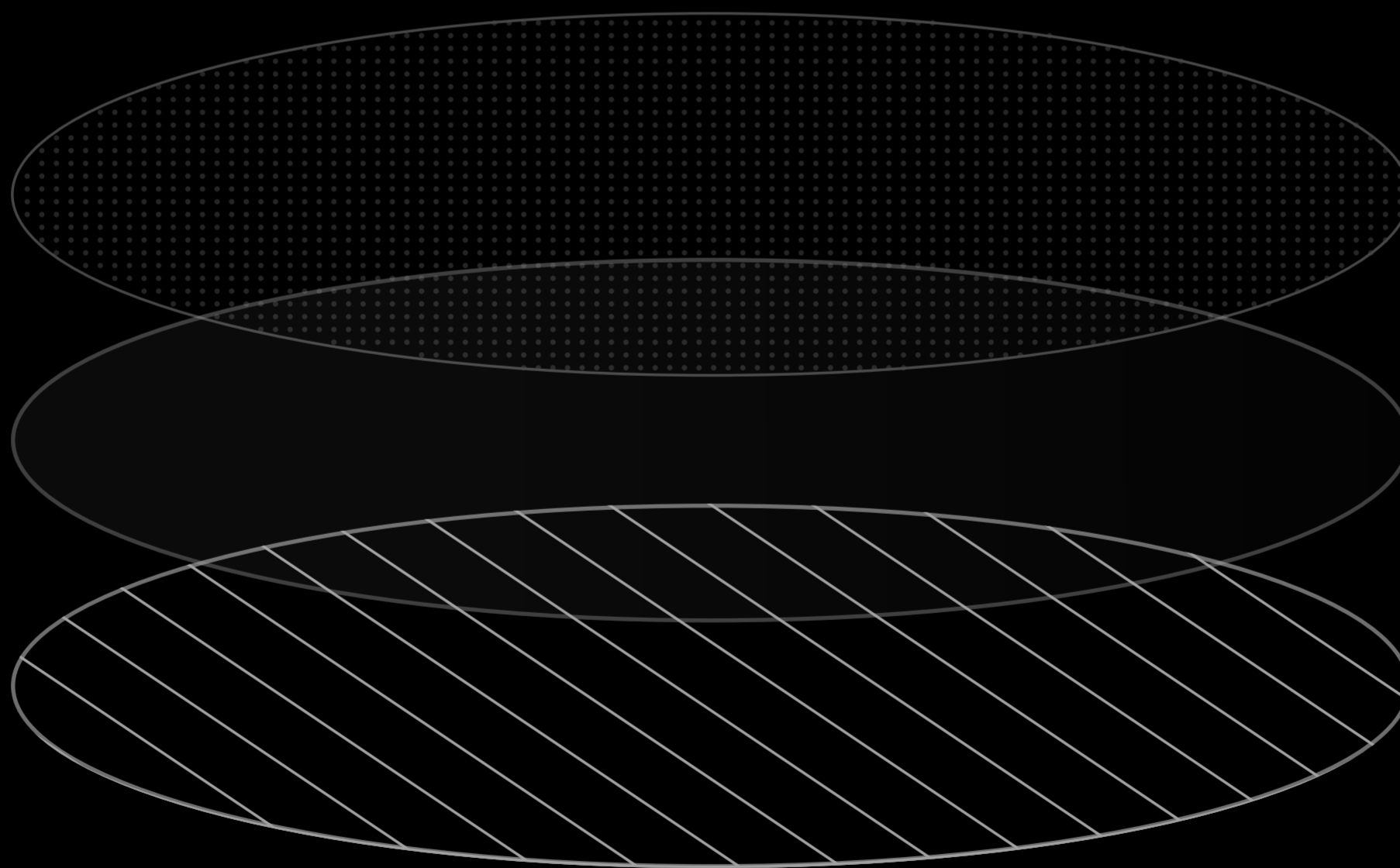


03

Finally, we bring in external, third-party auditors who review hundreds of additional examples to add an impartial and objective layer of quality assurance.

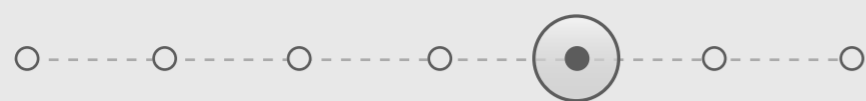
THIRD-PARTY ADJUDICATION

Independent clinical experts conduct large-scale evaluations to confirm unbiased, enterprise-grade performance



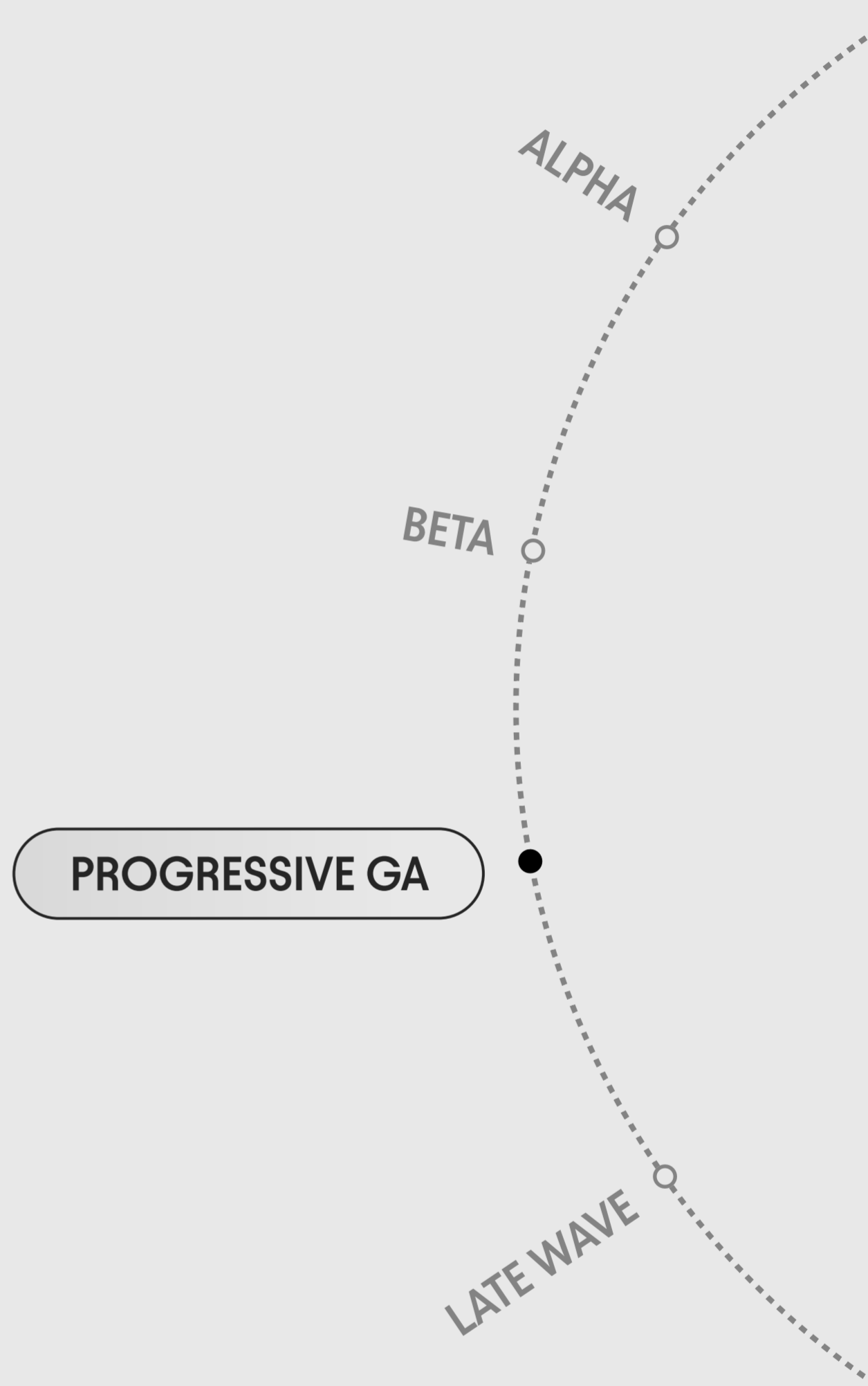


Only when a model clears every layer do we consider deploying. This multi-layer validation ensures that specialty model updates are not just different, but demonstrably better.



INTENTIONAL ROLLOUT

05







Intentional Rollout: The “Early Wave” Strategy

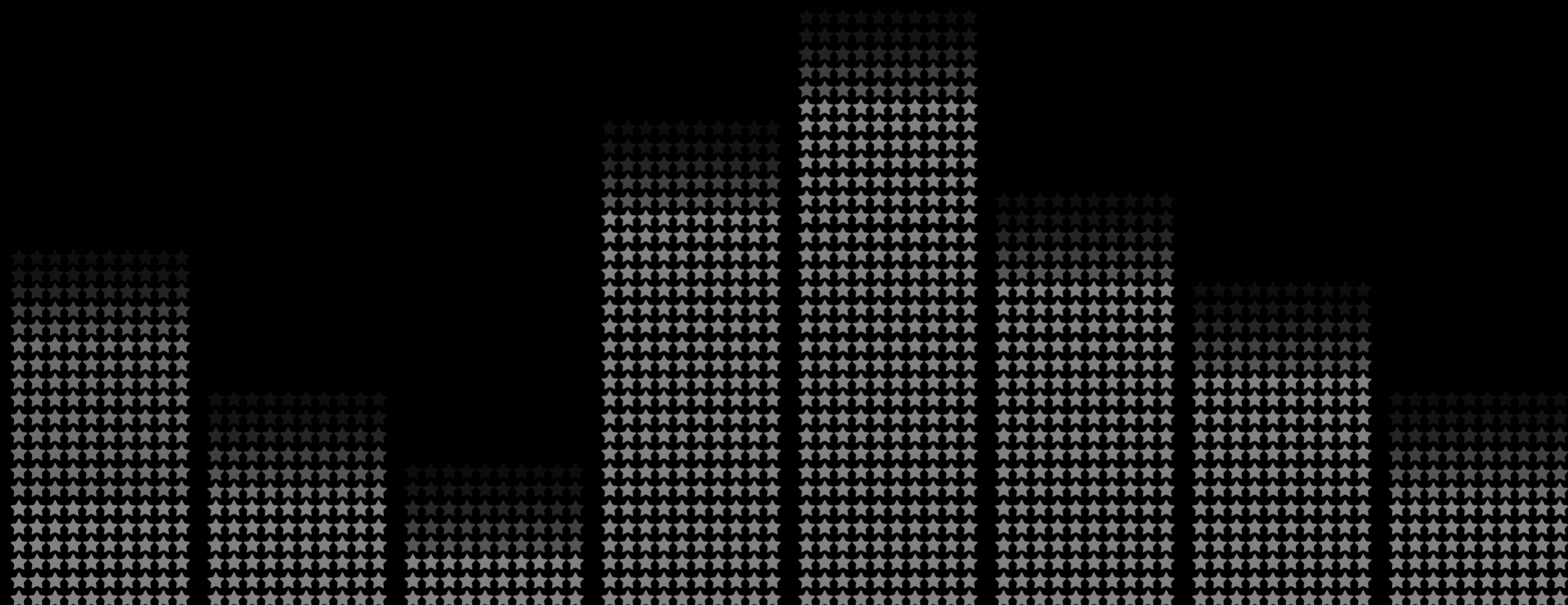
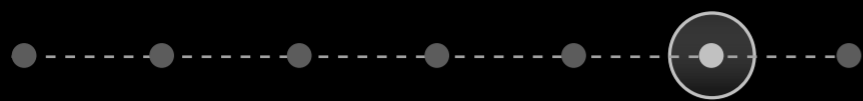
We don't flip a switch, we turn a dial.



At Abridge, our rollout process has four phases, each with defined audiences, expanding scope, and clear goals.

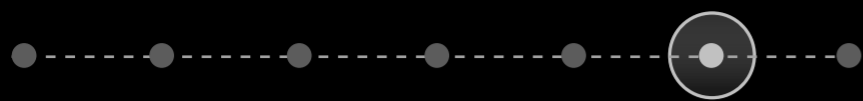
The table below shows how specialty model updates move through this strategy: progressing from Alpha to Beta, into staged General Availability with randomized clinician cohorts, and finally to 100% GA across our most complex partner environments.

 PHASE	 AUDIENCE	 SCOPE	 GOAL
ALPHA	Internal & External Auditors	Static Evaluations	Data Analysis & Benchmarking
BETA	“Early Wave” Partner Orgs & Clinician Champions	9 Orgs [14% volume]	Real World Feedback from Power Users
PROGRESSIVE GA	Randomized Clinician Clusters	25% to 50% to 100%	Statistical Validation via A/B Testing
LATE WAVE	Complexity-Heavy Organizations	Post 100% GA	Final Deployment to the Most Complex Environments



Scaling to GA: Silent Excellence

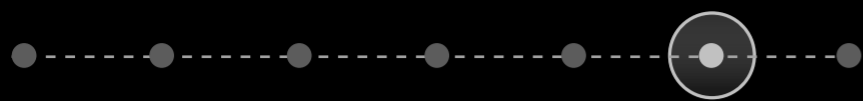
The transition from a "Beta" to "General Availability"
is defined by data, not dates.



General Availability isn't just a launch moment for us. It's a performance threshold. When a specialty model update is released, it's because it has already proven itself through multiple layers of quality and clinician review.

When we release an update to our models, our partners can trust that they have gone through rigorous review and validation to deliver improvement clinicians can experience. They will feel the difference because their notes will be all that much more aligned with their specialities.

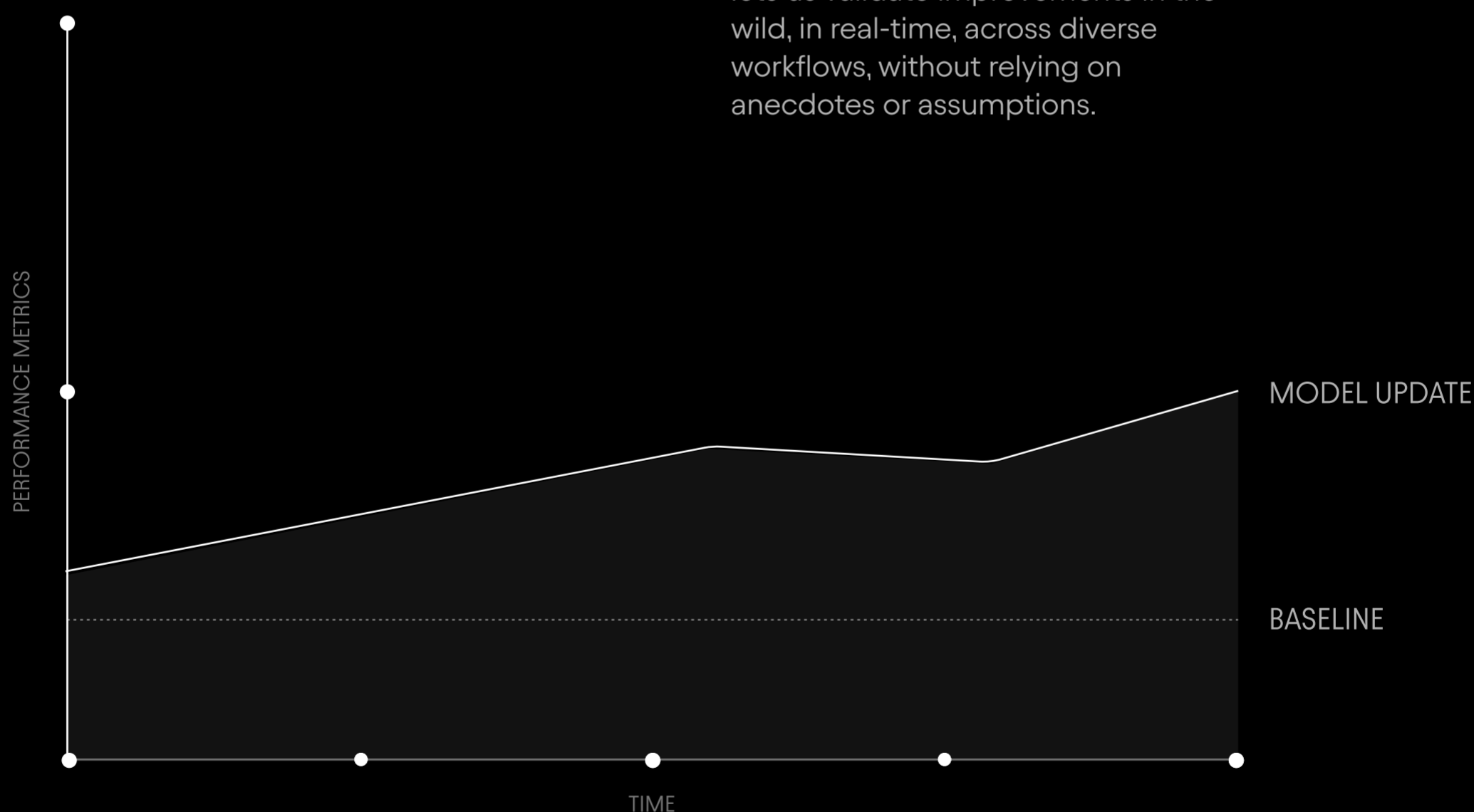
Reba Schenk
VP, Partner Experience
Abridge

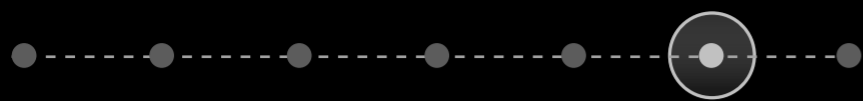


The progression to GA is governed by continuous measurement, not milestones.

Every specialty model can be observed in our live A/B testing dashboard where we monitor real-time star ratings alongside effort-reduction signals: how much clinicians edit, rewrite, or restructure notes.

When a specialty model consistently outperforms the baseline across these metrics, it earns its way into GA. This dashboard-driven approach lets us validate improvements in the wild, in real-time, across diverse workflows, without relying on anecdotes or assumptions.





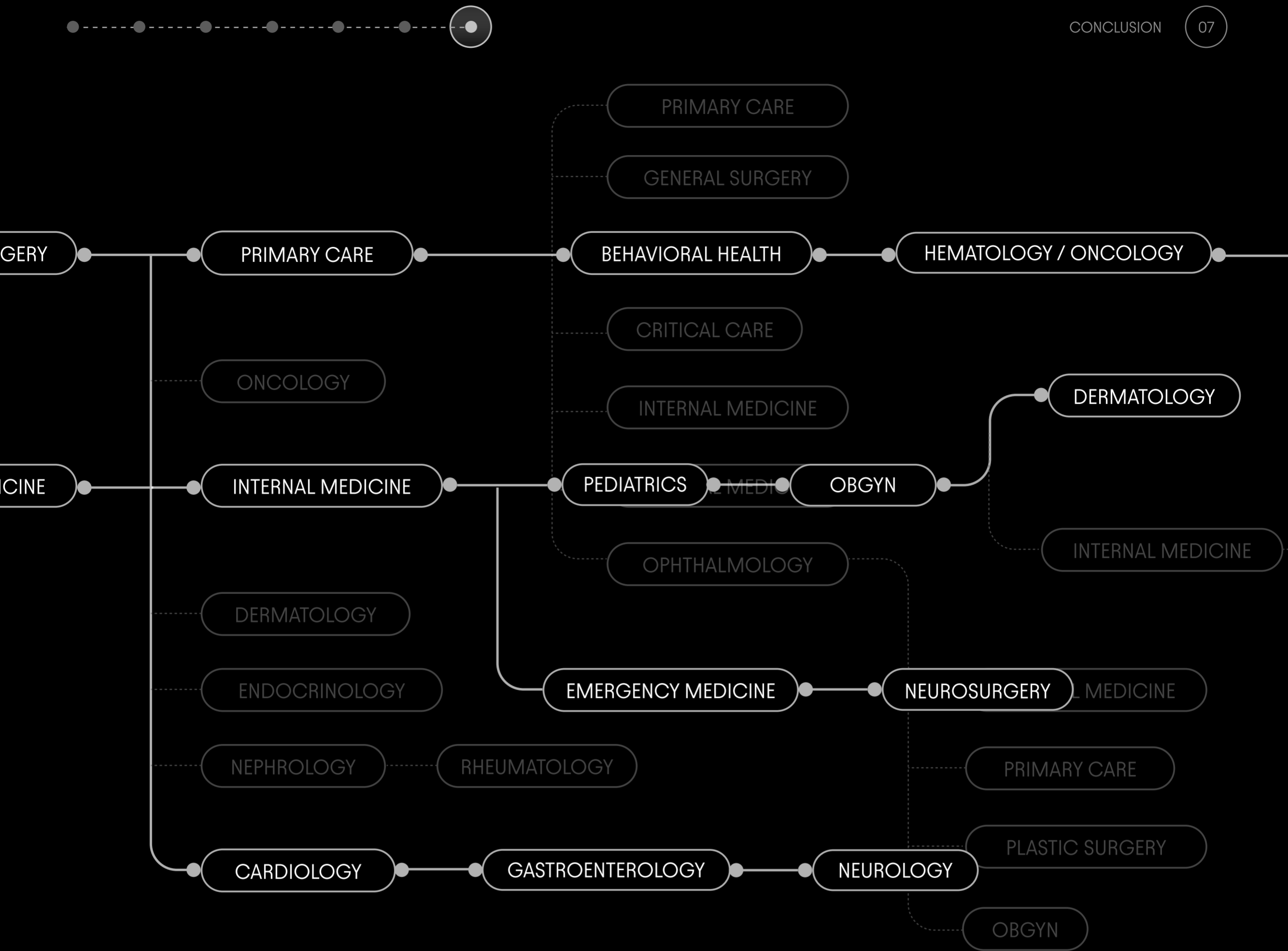
When models do move to GA, they do so seamlessly, deployed in the background without pop-ups, retraining, or workflow changes. From the clinician's perspective, notes simply read better, require fewer edits, and feel more aligned with their thinking, literally overnight.

At the same time, lightweight feedback loops remain active. One-click surveys and simple thumbs up/down signals help us capture feedback early, whether that's an Orthopedic surgeon preferring a "more concise" HPI or an Emergency Medicine clinician requesting a more tightly structured, problem-first assessment.

When sharing more nuanced or detailed feedback, clinicians leave comments directly in the app. At Abridge, feedback is foundational to the work we do. Feedback is our oxygen. Without it, continuous advancement would not be possible.

"Did Abridge get an update? Notes were good but now are 100% better. Kudos!"

Abridge Partner Clinician



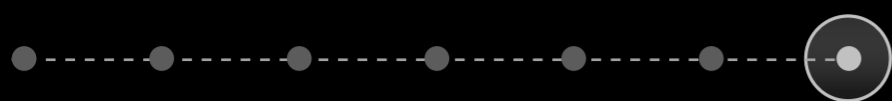
Conclusion: The Future of Iterative Intelligence

Specialty intelligence is a living system, not a product update.



By the end of the December 2025 sprint, Abridge moved several new specialty model updates into General Availability, including Hematology-Oncology, Gastroenterology, and the full surgical suite.

But more important than any number of updates is the framework behind it: a repeatable process to deliver deeply specialized intelligence without disrupting care. This is what allows us to scale across every corner of medicine while holding quality to an enterprise-grade standard. And the impact is measurable.



In some specialties, redundant language like “history of” in the HPI dropped from

46.8%

2.5%

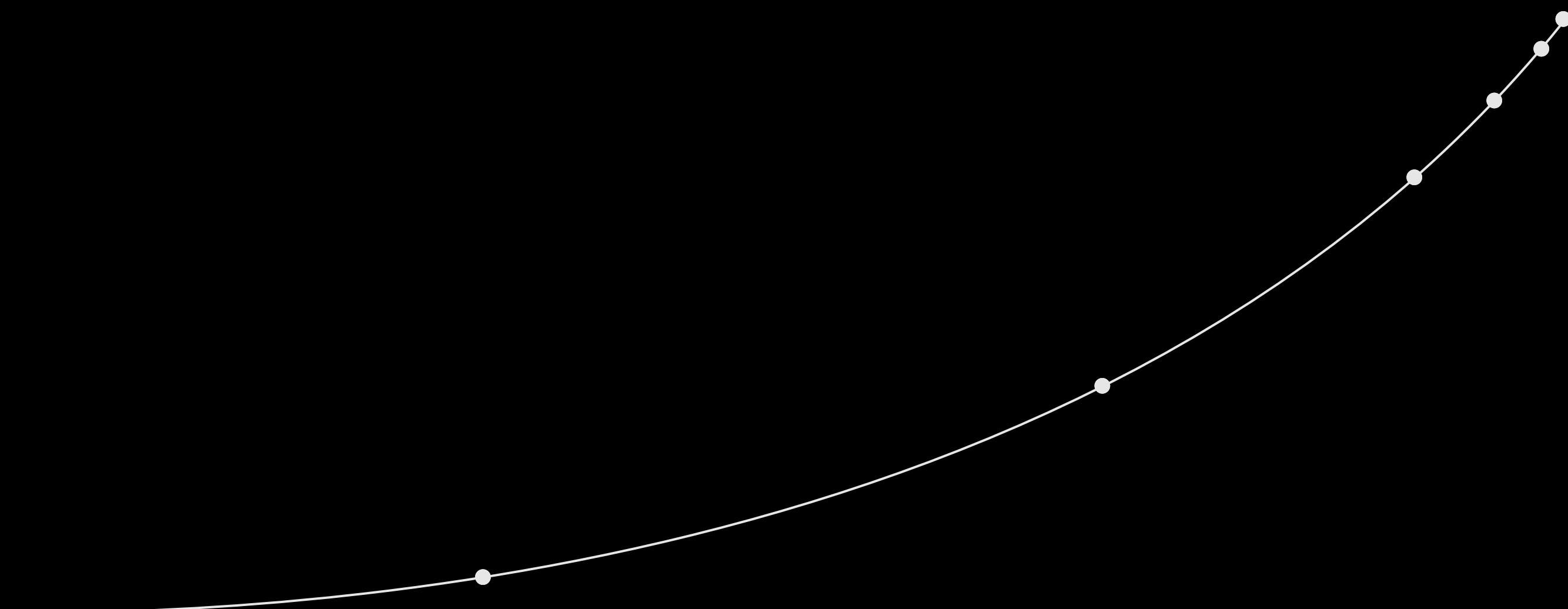
reflecting higher quality, “cleaner” notes.

These improvements are grounded in real-world validation from our Champion Network of hundreds of clinicians across specialties who provide the clinical, on-the-ground insights that help guide our work. Paired with focused Specialty sprints, this iteration framework compresses months of development and validation efforts into just weeks.



TESTING

04



The result is not just faster development, but rather a durable, reliable system built to improve our models. Delivering updates with measurable improvements into the hands of clinicians who know their care won't be disrupted? That's what it means to innovate at the speed of trust.