Rockin' HIT Sales

Episode Transcript

Episode: Why Clinical AI Dies in Pilot: From Impressive Demo to Hospital-Ready System
Guest: Sam Morhaim, CEO, Vantage IO
Release Date: May 20, 2026
Episode Link: Web Page Episode Link
Note: Transcript edited lightly for clarity and readability.

—

David Hacker (00:13)

This is Rockin' HIT Sales, a quick podcast built around the buyer-side realities of healthcare. I'm David Hacker, your host and founder of Elevate HIT Sales. In each episode, I sit down with health system and Health IT leaders to unpack what actually gets initiatives approved, implemented, and scaled. Today, I'm joined by Sam Morhaim, CEO of Vantage IO, for a practical conversation on why clinical AI tools can look so impressive in a demo, but still fail when they hit hospital review, validation, workflow, compliance, and real-world deployment. We also discuss why RAG and evidence grounding are becoming such major issues in clinical AI, and what founders, CTOs, product leaders, and go-to-market teams need to understand before claiming their AI solution is ready for provider environments.

David Hacker (01:21)

Sam, welcome to Rockin' HIT Sales. I really appreciate you joining me today to talk about why clinical AI tools can look impressive in a demo, but sometimes fail when they hit real-world deployment. This is an extraordinarily timely topic for HealthTech founders, CTOs, product leaders, and the GTM teams trying to understand what it really takes to move from 'works on GPT' to something that is ready for prime time.

Sam Morhaim (01:54)

Yes, thank you so much. Thank you for having me.

David Hacker (01:57)

You have been writing about why clinical AI dies in pilot. When you say that, where do you most often see it break? Is it in the technology, the workflow, the evidence, or the review process?

Sam Morhaim (02:13)

That is a good question. Almost never is it the model. Usually it is the layer around it. The top failure points typically are the grounding of the evidence, the audit trails, and the failure handling - basically handling those edge cases and all the things people did not think about when they were testing in pristine testing environments, even when they use synthetic data. That is something that is not prepared correctly, and that is where we see most of those failures.

David Hacker (02:45)

A lot of HealthTech companies today can quickly build a very impressive AI demo. From your perspective, what separates those nice, shiny demos from something that is actually ready for clinical review?

Sam Morhaim (03:01)

AI has unleashed the power for everyone to build solutions, and that is great. But at the same time, there are a lot of complexities under the hood that people do not understand. Even if you are aware of HIPAA, GDPR, and other compliance and security requirements, building something that works as a demo is one thing. Adding all the constraints and layers that make sure it is safe, efficient, and useful for doctors or ultimate users is a completely different game.

Sam Morhaim (03:01)

A lot of people underestimate that and often leave it for later stages. That is where many issues start building up. Because you leave it for later and do not think about it early on, at some point you may have to do something as small as a refactor or as big as a full rewrite, as we have seen with some of the clients we have helped in the last few months.

David Hacker (04:07)

Is that where you see the biggest misconception digital teams have between the gap of passing clinical risk assessment and working on GPT?

Sam Morhaim (04:20)

Yes. There is a big gap in expectations when you use ChatGPT, Grok, or any of those tools directly. They have so many additional layers under the hood and additional piping. To give you an example, with a client we worked with to ingest labs and process them into patient charts, they were testing directly with the UI of ChatGPT, Claude, and other tools. It worked really fast and appeared to work really well.

Sam Morhaim (04:20)

When you use the API, a completely different story happens. Starting from the base system prompt all the way to the architecture behind how the file is chunked, indexed, and processed, there are many things you do not think about when you are building a prototype or even vibe coding. Even if you have a proper team building it, sometimes they do not understand those gaps. You immediately get that expectation difference between what you think is happening and what is actually happening in your own application.

David Hacker (05:30)

Is model accuracy alone enough to make a clinical AI tool trustworthy, or does it have to go deeper?

Sam Morhaim (05:40)

No, it goes much deeper. I will use an example I often use. The models are very well trained on a lot of topics, and they are very good at sounding smart. I am not saying they are not smart. They are helping us get where we are today after two or three years of using them, so they are very useful.

Sam Morhaim (05:40)

But it is like asking someone questions about all sorts of topics - weather, for example. They may have basic knowledge and can even sound smart when answering. But only when they actually sit down and read everything there is to know about weather, current patterns, and different types of clouds do they become expert enough to make sense of what they are saying and know when they are failing.

Sam Morhaim (05:40)

That is one of the biggest differences between a proper system and a prototype: the system's ability to know that it is failing or that it does not accurately know the answer, and to be honest about it. That is one of the top issues we find with these models.

David Hacker (06:55)

With your level of expertise, what are the warning signs that a clinical AI product was built for demo impact rather than real-world deployment?

Sam Morhaim (07:13)

The first thing you can assess is the confidence or certainty it shows when it makes assumptions or statements. If there is no traceability to the source of the information, that is number one. Another signal is whether the system can say, 'This is a weak link,' or 'There is not enough evidence for this, but I found this, and that is why I am surfacing it.' That self-awareness or self-honesty is important.

Sam Morhaim (07:13)

Citation sourcing and auditability of the answers are also key. You need to be able to go back under the hood and say, 'This answer came from this endpoint in the RAG pipeline,' and have all that traceability. If that is missing, it is a big telltale sign that it is just a prototype and not a real production system that can withstand the scrutiny of enterprise procurement.

David Hacker (08:11)

Where do security, compliance, and clinical teams on the facility side tend to find problems that product teams did not anticipate?

Sam Morhaim (08:25)

There are two main areas of concern from the buyer side. One is change management. Because models and technology are evolving so fast, the old model was that you would certify a version, go through a few months of procurement, and that version was locked. Now the model can evolve, and parts of the code can evolve almost daily. We see companies pushing code multiple times per day.

Sam Morhaim (08:25)

The underlying model and logic might change, and the hospital or buyer may be unaware of it. The consequences may not be immediately visible and can go under the radar. The second area is edge cases. There will always be an extreme case, or even a not-so-extreme case, that is outside the happy path where things might break. If there are not enough safeguards built into the tool, the hospital security team or pilot supervisors will not have enough information. They need telemetry and observability tools that allow them to prevent these issues or at least observe them before they reach patients or users directly.

David Hacker (09:54)

How should companies think about retrieval-augmented generation, or RAG, compliance and evidence grounding in a clinical environment?

Sam Morhaim (10:06)

That is a big one because most teams are not fully understanding RAG yet. At the core, there are a lot of security considerations. This is a new component. Just like when we added databases or file storage, now we are adding RAG. RAG can, at some point, cross-pollinate or spread data that you did not intend to be seen.

Sam Morhaim (10:06)

The bot or LLM can query the RAG and expose data from two separate patients, even if you index it correctly and do many things right. There are additional layers and techniques you can apply to help ensure that does not happen. But probably the biggest PHI leakage risk we have seen happens at the RAG layer. Teams may take real patient data and use it to train models or as foundational data ingested into the RAG. Now real PHI sits on those RAG servers, serving data to the LLM, where it could potentially be reverse engineered to identify a patient, even if de-identified, or simply expose information that was not meant to be shown or learned by the LLM.

David Hacker (11:31)

In your writing, you have outlined six architectural layers that separate a clinical-grade system from a demo. At a high level, can you walk us through those layers and why they matter?

Sam Morhaim (11:48)

One of the most important is traceability: being able to know exactly at what point information came to your LLM and where it came from. That is one of the first things. You also need to know that similarity in RAG is not the same as relevance. That is another major issue.

Sam Morhaim (11:48)

Another important layer is logging. Not just traceability and knowing where information came from, but how you log the data that a prompt went in, what system prompt came with it, and all the artifacts you sent to the LLM. On the way back, you need to know what the LLM returned, and you need the ability to safely store that and safely retrieve it if you need to reconstruct an answer. Even for internal improvement, that is critical, and many teams do not have it done well or securely enough.

Sam Morhaim (11:48)

Another layer is the ability to speculate and be honest about it. LLMs, as we have seen with multi-agent systems, like to speculate with one another. The system needs to be able to say, 'I think with high certainty this is the right answer,' or 'With a lower degree of certainty, this might be the answer,' and state that transparently. Those are some of the top layers and issues that typically break.

David Hacker (13:41)

Failure modes are probably the least flashy layer of AI, but also one of the most important. What should an AI system do when it either does not know or should not answer?

Sam Morhaim (14:01)

Number one, it should be honest about it. There is an evidence-fit issue where LLMs do not like to say, 'I do not know.' At the core, we need to give the system the ability to say, 'I do not know,' or 'My certainty level is very low,' and surface that.

Sam Morhaim (14:01)

It can look different in different systems. In some platforms we have developed, the system says, 'The evidence is not strong enough, but here is what we found,' so that it is honest about it. Those are two of the most important safeguards to add.

David Hacker (14:49)

For a digital health team preparing to sell into a facility, what should they have ready before they even enter a clinical AI review process?

Sam Morhaim (15:03)

At the core, they need to be able to answer the primary question: where does PHI data live and where does it go? Today, with so many endpoints and models, the data can live and flow in many different directions.

Sam Morhaim (15:03)

Even if you have a business associate agreement, data should not just flow blindly to one of the LLMs. Even if you think you have it on AWS Bedrock with a BAA, data should not flow there freely. There need to be safeguards. Your ability to map where your data lives, how it is stored, and where it flows at all times will help you answer multiple questions on those questionnaires.

David Hacker (15:52)

How can AI companies talk about their capability without overclaiming or creating risk for themselves during the review process?

Sam Morhaim (16:04)

At the core, it is about being honest. We have seen people claim they have zero hallucinations, and we need to be honest that LLMs and AI are, by nature, non-deterministic systems. If you are familiar with the butterfly effect or chaos theory, a small change or variance in one part of a non-deterministic system can end up producing a very different output. That honesty and self-awareness are important.

Sam Morhaim (16:04)

The other piece is knowing that you cannot test what you do not know. You do not know what you do not know. You have to keep pushing the boundaries of what needs to be tested - not just with synthetic data and real data, but by asking what edge cases or combinations of things you have not tried yet. You will never get to 100 percent on a non-deterministic system, but the closer you can get to that mark, the more confidence you can build, and the more successful pilots and implementations can be.

David Hacker (17:19)

Can you offer guidance to teams on documentation, evidence, or architecture stories that will help a facility trust that the company actually understands clinical risk?

Sam Morhaim (17:35)

Currently, teams are doing a pretty good job, but they need to go beyond checking the box. We live in a world that is moving so fast that having a questionnaire with 100 checkboxes and just passing it to the next department is not going to cut it. They need to roll up their sleeves, get into the system, understand how it works, and understand where PHI moves through the system.

Sam Morhaim (17:35)

They need to try not only to make sure it works, but to break it and see where it does not behave as expected. That would alleviate a lot of the pain we have seen. You need to go past the checkbox mentality.

David Hacker (18:16)

I think we are past the question of if. Now it is a question of how. How should go-to-market teams adjust their messaging when selling clinical AI into hospitals?

Sam Morhaim (18:29)

Right now, the marketplace is so crowded. Everyone has been unleashed, and now they are creating solutions left and right. There needs to be a differentiator. Everybody has stamped AI on their name, or claims to be AI-native. But just adding a chatbot to your product does not make it AI-native.

Sam Morhaim (18:29)

They need to adjust the messaging to stand outside and above the crowd. They need to find the real gaps because now everybody is AI. What is your true differentiator? What makes you different and sets you apart from all the others, not just on cost? What really matters is whether this will make a big enough splash to justify the effort and go above and beyond what competitors are doing.

David Hacker (19:25)

Because of this tidal wave of AI and everyone putting AI onto their logo, is it fair to say that now, as part of go-to-market readiness, architecture needs to be an important component of the sales motion conversation?

Sam Morhaim (19:45)

Absolutely. As regulation catches up, including as HIPAA adds requirements around the safe and efficient use of AI, architecture is going to be at the center. We will have additional tools from the buyer side to know that what you are claiming is actually true. Before, it might have taken a due diligence team to inspect that. Now it may take only a few commands to know for sure. Transparency and preparation of the architecture - making sure it is correct and up to the standards of what you are selling - will become crucial.

David Hacker (20:27)

That brings us to the final two questions, what I refer to as my lightning wrap. The first question is: what is one thing clinical AI companies should stop saying in front of clinical AI reviewers?

Sam Morhaim (20:51)

Stop saying hallucination-free, because it is a fallacy. We need to be aware of that and surface it instead of trying to hide or negate it.

David Hacker (21:05)

What is the one thing every AI company should build or document before they start claiming their AI is clinically ready?

Sam Morhaim (21:20)

I do not think there is a single thing. It is a combination of all the right things done in the right sequence. Having a trusted partner or a trusted set of advisors who have seen this before and have done it before to guide you through that is very important. That is why teams need to do it as early as possible in their process.

David Hacker (21:46)

Sam, thank you so much for joining Rockin' HIT Sales. This was a great conversation on why clinical AI readiness is not just a technical issue. It is trust, evidence, workflow, and go-to-market. I am 100 percent positive our listeners will walk away with a much clearer understanding of what reviewers are really looking for and what founders need to build and document before claiming their AI is ready for clinical environments. Thank you.

Sam Morhaim (22:16)

Thank you. You are welcome.

David Hacker (22:18)

As I do with every episode, let's take a quick look at my three brief takeaways from my conversation with Sam.

David Hacker (22:18)

Number one: clinical AI rarely fails because of the model alone. The weak points are often around the model: grounding, audit trails, failure handling, workflow debt, PHI movement, and the safeguards needed for real-world clinical environments.

David Hacker (22:18)

Number two: RAG is becoming a major clinical AI risk area. Retrieval-augmented generation can be very powerful, but if teams do not understand how the data is indexed, retrieved, separated, and exposed, RAG can create serious PHI and evidence-grounding problems.

David Hacker (22:18)

Number three: architecture now has to be part of go-to-market readiness. Clinical AI companies need more than a strong demo. They need to explain how the system works, where the data flows, how evidence is grounded, what happens when the system is uncertain, and why hospital reviewers should trust it.

David Hacker (22:18)

Have questions, comments, or guest ideas? Email me at podcast@elevate-hit-sales.com. Music is courtesy of Pennsylvania's finest, my friends, The Badlees. And since I am supporting Bachman-Turner Overdrive today, remember: if your clinical AI only looks good in the demo, you ain't seen nothing yet. To get hospital-ready, you need evidence grounding, audit trails, safe failure modes, and a system that can truly start taking care of business.