HALO LAB · WORKSHOP

Why Healthcare AI Fails After the Prototype Stage

A working session on why healthcare AI stalls between prototype and production, and what the teams who actually ship do differently.

PUBLISHED 15 APRIL 2026 • 49 MIN • WORKSHOP

HOST
Sergiy Sumnikov
General Manager, Healthtech · Halo Lab

GUEST
Dmytro Loposhansky
AI Technical Lead · Seattle Children's Hospital

KEY TAKEAWAYS

The practical throughline of the session, distilled into moves a team can act on.

01 Validate on real-world data before you trust a model.

Models tuned on clean public datasets break the moment real EHR data arrives, because of missingness and site-specific field mappings. Go to great lengths to test on real data first.

02 Design for site-by-site variation from day one.

Every health system runs a different vendor stack (Philips, GE Healthcare, Dräger, Spacelabs). Build adapters that normalize inputs into your own format instead of asking the site to adapt to you.

03 Make the model retrainable on the local population.

The Epic sepsis model degraded because accuracy shifted with each population. Shipping a model customers can retrain on their own data is what recovered its quality.

04 Put clinical expertise on the team early.

Good-enough validation needs a clinical champion who can guide customization. In healthcare there is no one-size-fits-all, so the product that adapts is the one that wins.

05 Have the compliance pack ready before you approach IT.

Prepare the BAA, MSA, SLA, and SOC 2, and expect a consensus-based AI review board. Large institutions procure over months, not weeks.

06 Budget for workflow, not just research and engineering.

Projects stall on people and workflows, not technology. Fund change management, and make feedback and change visible so clinicians feel in control.

07 Prove ROI with a simple before-and-after story.

Lead with who had the problem, what changed, and what it is worth, the way the Cleveland Clinic referral example freed several full-time coordinators by automating most of the workflow.

IN THIS SESSION

What decides whether a healthcare model makes it from a promising demo into daily clinical use.

00 **Introduction**

Backgrounds, the four themes, and how the session runs.

01 **Validation**

Getting real-world data, spotting models that won't generalize, and what "good enough" looks like.

02 **Integration**

Compliance and procurement, demo-ready versus production-ready, and where projects stall.

03 **Trust & Safety**

Losing and keeping clinician trust, explainability, and the design choices that decide adoption.

04 **Return on Investment**

Who decides, which metrics matter, and how to prove value before full deployment.

About this transcript. Edited lightly for reading: filler words and false starts removed, and a few clear mishearings corrected. The speakers' meaning is unchanged, and nothing has been added. Pull quotes are drawn verbatim from the conversation.

00 Introduction

Backgrounds, the four themes of the session, and how the hour will run.

Sergiy Sumnikov
HOST

Hello, everyone, and thank you for being punctual. My name is Sergiy Sumnikov. I work at Halo Lab, and today I have the pleasure of welcoming Dmytro Loposhansky, AI Technical Lead at Seattle Children's Hospital. Hi Dmytro, how are you doing?

Dmytro Loposhansky
GUEST

Hi Sergiy. Thank you, thank you. I'm glad to be here.

Sergiy Sumnikov
HOST

Today we'll cover a lot of practical, hands-on material: recommendations, tips, and so on. The topic, as you know, is why healthcare AI fails after the prototype stage. Let's start with a bit of background. Could you tell us more about yourself, especially where it concerns AI and hospitals?

Dmytro Loposhansky
GUEST

I'm currently AI Technical Lead at Seattle Children's Hospital, where I oversee the development of internally built AI tools, solutions, and machine-learning systems. My role is to find and identify inefficiencies and problems within the health system, propose AI solutions, and then actually go and implement them. So I run the full cycle, from ideation to production deployment, and I've done that now for multiple departments across clinical decision support, administrative systems, and general automation.

Before Seattle Children's I worked at SickKids in Canada, building the AI infrastructure for the hospital. My training is in computer science at the University of Oxford, and I've since worked across several domains, but healthcare is the one that's closest and dearest to me. I think it's the domain that stands to gain the most from AI, and there's still very little implementation done. My role is to bridge that gap and move exciting research into real healthcare settings, which sounds easier than it is in practice. I'd be happy to share some of those learnings today.

Sergiy Sumnikov
HOST

Great. Could you also give a few examples of the AI products you've worked on?

Dmytro Loposhansky
GUEST

Absolutely. Right now my focus is on three clinical decision-support systems. One is an electrocardiogram foundation model. It works across two modalities, signal and text, so it can read a 12-lead EKG, determine the likely diagnosis, and provide a clinical-grade rationale for why it thinks those statements are correct.

That one model can scale massively. Seattle Children's serves patients from five states across the Pacific Northwest, and some of those states don't even have electrophysiologists available to read EKGs. A system like this could scale well beyond our own health system, across the region and hopefully wider. Alongside that, I'm working on models that detect hidden cardiovascular states and change points in signals, and on automation on the administrative side, which we all know is a huge burden in healthcare. Some of these aren't publicly available yet, but I'll bring in examples as we go.

Sergiy Sumnikov
HOST

Thanks. Just to mention the format: at the end we'll have a Q&A where attendees can ask questions. If you have more detailed or specific questions that need deeper context, feel free to send them to me by email, and I'll forward them to Dmytro in a structured form.

We have four subtopics today: validation, integration, trust and safety, and ROI. I'll post them in the chat, and I'd ask attendees to drop in the number of the subtopic that feels most relevant to them, so we can allocate our time well. This is also a good moment to mention today's sponsor, Halo Lab, where I work. We help AI product builders improve adoption and outcomes through better UX design. Let's jump straight into the first section.

01 Validation

Getting real-world data, the early signals that a model won't generalize, and what counts as "good enough" before moving on.

Sergiy Sumnikov
HOST

Dmytro, if someone here has a model that performs well on test data, what's the first real-world validation step they should take?

Dmytro Loposhansky
GUEST

I've had a lot of experience with exactly this framing. At SickKids, our department collaborated closely with a group of researchers at the University of Toronto, where graduate students were building models on publicly available datasets, some of them research-only.

None of that data was real-world quality, which is very often far more degraded than the polished, pre-processed datasets you find online.

So the number-one thing to do is get real-world data, because that's the essence of validating whether the approach actually works. My job was to take the models those researchers developed and move them into a production setting inside a health system. And the moment real data was fed in, the models immediately broke. There's missingness. The model expects the data to always be complete, and sometimes we only have twenty percent of what it expects, so it has to know how to handle that. If you look at the electronic health record and its specific fields, every health system maps and records attributes differently, and all of that has to be accounted for in production.

So go to great lengths to get real-world data. Some companies will run a free trial or a free period for their first client just to get real examples. There are also accelerators founded by health systems that can provide real-world data to startups. Whatever it takes, try your system out in the real world.

Sergiy Sumnikov
HOST

What can be early signals that a model won't generalize once it's inside a real hospital?

Dmytro Loposhansky
GUEST

Every health system is slightly different. I've heard the expression that each one is like a snowflake; you won't find two identical ones. It depends heavily on scale, too. Academic medical institutions with many thousands of staff have their own practices, their own IT systems, their own workflows, and it all differs.

The number-one red flag I look for is a lack of adaptability. Here's a concrete example. When I was building the AI infrastructure for real-time physiological data collection, we built four or five different adapters. The infrastructure was the main piece, but every health system uses a different vendor for data acquisition, Philips, GE Healthcare, Dräger, Spacelabs monitors, and others. We can't ask a health system that has invested millions in its infrastructure to adapt to our technical solution. We have to adapt to the real-world diversity.

So each adapter was a software component matched to a specific device manufacturer, and every one unified the incoming data into our own common format, which was then used throughout the infrastructure. Without that, we could never have scaled. The system

is now live on almost every continent, in Australia, Canada, the UK, Israel, and a few other places. Every country has its own stack, and I'm very glad it scaled the way it did.

We can't ask a health system that has invested millions in its infrastructure to adapt to our solution. We have to adapt to the real-world diversity.

DMYTRO LOPOSHANSKY

Sergiy Sumnikov
HOST

Let's put our product-management hats on. Have you seen technically strong models fail because they weren't designed around real clinical workflows from the start?

Dmytro Loposhansky
GUEST

The one that comes to mind is the Epic sepsis model, which many people will have heard about. It was developed by Epic Systems, the largest EHR company, and they carry this reputation that whatever they build must be best-in-class simply because they're the biggest player. The real-world story turned out to be quite different. When the model was rolled out and institutions adopted it, the originally claimed accuracy of around eighty percent turned out to be closer to fifty to seventy in practice, and at some institutions the sepsis predictions were close to random.

Epic has since released a new version, but it still produces different results at each site. The key insight is that if you're building AI tools on real patient data, every institution has its own population. The population here in the Pacific Northwest is different from one in Europe. People genuinely differ, ages differ, and all of it affects quality. If you can't evaluate the model across different patient populations and make it retrainable, you're stuck. One of the key fixes was allowing retraining on local populations: they shipped the model, each customer retrained it on their own data to fit their population, and quality improved. It's a famous example, and it illustrates the concept well.

Sergiy Sumnikov
HOST

I'd like to close this first section with a question. What does good-enough validation look like before moving forward? Are there criteria or aspects you'd highlight?

Dmytro Loposhansky
GUEST

First, you need to be adaptable and have the components in place to handle the heterogeneity you'll meet in the real world. Second, and just as important, you need the right expertise on the team to address whatever comes up. Whatever you're building, it's a good idea to have someone with clinical expertise, a champion, who can guide you through the steps needed to customize the model.

In healthcare there's almost no such thing as one-size-fits-all. The winner is the company or product that can adapt, and if you can validate that during your demo stage or first iteration, that's probably the main thing to focus on.

In healthcare there is almost no such thing as one-size-fits-all. The winner is the product that can adapt.

DMYTRO LOPOSHANSKY

02 Integration

What to prepare before approaching hospital IT, the gap between a demo and a production system, and where projects stall.

Sergiy Sumnikov
HOST

Let's move to integration. What are the key things a team should prepare before even approaching hospital IT?

Dmytro Loposhansky
GUEST

Number one is compliance. You have to be ready to provide the business associate agreement, the MSA, the SLA, all of those core documents, from the get-go. Certifications matter too. We always ask for SOC 2 or other security documentation as part of the IT intake process.

It's also important to understand that when large institutions adopt software, they don't just purchase, they procure. That's a long, multi-month, sometimes year-long process where many people and groups evaluate the tool: legal, security, IT, and the clinical or administrative side, depending on where you're integrating.

And if you're building in AI specifically, be ready for a review board. Many hospitals across the US are setting up AI review committees

made up of key leaders from around the organization. They assess your solution against ethical criteria and against what each of their departments cares about. Decisions are consensus-based. There's no single person who says "I'll buy this for everyone," because no one person holds the expertise of an entire institution. Be ready for that.

Sergiy Sumnikov
HOST

How different is a demo-ready system from a production-ready one in healthcare?

Dmytro Loposhansky
GUEST

One aspect we haven't touched on yet is the financial model. I was reading recently that there are more than a thousand FDA-cleared, AI-based software-as-a-medical-device products, but not many of them are actually reimbursed by insurance. So even after you've built the tool, implemented it, and gone through the long regulatory-approval process, it still isn't fully ready. There's still the question of whether you can make the economics work inside the complicated world of healthcare.

That's where things get stuck: the deployment, and proving the tool delivers specific, quantifiable value and ROI, and building that into reimbursement workflows. So a production system is one that has the economics figured out. Often it's less about the product itself and more about smart business thinking, working out where the value is and quantifying it, so the system works as a whole.

A production system is the one that has the economics figured out.

DMYTRO LOPOSHANSKY

Sergiy Sumnikov
HOST

Where do projects most often stall? Is it IT, security, compliance, or workflows?

Dmytro Loposhansky
GUEST

That's a tough one, because every stage has its own barriers and it shifts with the company's maturity. But the workflows are where I see the most friction. You can build the tool and get it through IT and security review; that part is completely doable, and there's a clear checklist to complete. The harder part is people.

It's the people problems that are the most difficult. The technical things aren't as hard, in comparison, as dealing with humans. Everything to do with workflows becomes difficult, because now you're dealing with the people on the floor, your end users, who are already accustomed to how they do things. How do you drive change management? That's genuinely hard.

It's usually solved through a champion or specific rules. One example I heard: a health system wanted clinicians to adopt a new procedure, so they set up a scoreboard that ranked people by whether they complied, and it was updated on a set day. Nobody wanted to be at the bottom of the list, so adoption climbed. It was a gamified way to promote it, and I found it fascinating how human psychology works here. You don't want to be pushy or force your tool on people, but even when something is genuinely useful, people need time to adapt.

It's the people problems that are the most difficult. The technical things aren't as hard, in comparison, as dealing with humans.

DMYTRO LOPOSHANSKY

Sergiy Sumnikov
HOST

You mentioned adoption, and it's an important term with many facets. Let me share a little of what we've seen with our own customers, where adoption failed or could be significantly improved.

A lot of it is about workflows. A new interface or a new login page is extra effort, and clinicians are already heavily loaded, so it becomes a blocker. Another thing we see, and I think you've experienced it too, is cognitive overload. A dashboard with twenty parameters and graphs is very hard to take in, especially in settings like the intensive care unit, where clinicians have only seconds to make important decisions. We solve that as a UX agency through better prioritization: first show what's truly urgent and relevant, in the right order, then use progressive disclosure so people can dive into the raw data when they need it.

One more, especially for teams that are very data-science-focused: don't organize the interface around your data source or ML output. Clinicians have their own mental models, usually rooted in physiology, body systems, and cause and effect. That's what's in

their heads. There's more I want to say here, but it leads naturally into our next section, so let's move to trust and safety.

03 Trust & Safety

The moment clinicians lose trust, the non-negotiables for safety, and the interface choices that decide whether a useful system gets used.

Sergiy Sumnikov
HOST

My first question in this block: from your experience, what's the first moment clinicians lose trust in an AI system?

Dmytro Loposhansky
GUEST

It might be obvious, but it's when the system makes mistakes, or when it contradicts a clinician's intuition. If your end user sees it getting things wrong, or behaving in a way that doesn't match their expectations, a chasm opens between what they want and what they're actually seeing.

You mentioned ICU workflows, and that's one of the areas I work in closely. We have a system called GE Healthcare MUSE for electrocardiology. Today's state of the art there is an automated, rule-based interpretation system: when an EKG is taken, embedded rules generate the statements. When I've spoken with our electrophysiologists, they tell me they never look at it, because it's never fully correct. They prefer to delete whatever the tool suggested and type everything from scratch, because they don't even want to be biased by looking at the model's output.

That shows how a tool meant to be useful can completely lose trust. It still runs, it's still there, but for actual experts reading these tests, it's not useful, and they skip it entirely. People judge these systems on expectations versus reality, which makes sense.

> *They prefer to delete whatever the tool suggested and type everything from scratch, because they don't even want to be biased by looking at it.*
>
> DMYTRO LOPOSHANSKY

Sergiy Sumnikov
HOST

What are the non-negotiables for making an AI system safe in a clinical environment?

Dmytro Loposhansky
GUEST

Let me continue that same example. The way we're trying to solve it is by building our own custom foundation model that reads the electrocardiograms. We didn't want to just mimic the old system, take the test in and push a diagnosis out. For the system to be safe in operation, we needed the explainability piece figured out. In general, the transparency of your AI tool is what makes the difference between "I trust this" and "I don't."

Concretely, for every statement the model makes, there's a rationale. If you use large language models today, they all have chain-of-thought reasoning: the model doesn't just spit out an answer, it works through internal thinking it can display, so you can trace how it reached its conclusion. We do the same thing here. We let the clinician inspect the reasoning trace, and if it matches their intuition and their thinking, the diagnosis is confirmed.

It closely mirrors how human clinicians learn. When we have fellows on service, they confirm a diagnosis by explaining their reasoning to a senior clinician. So in that sense we can mimic the human process to keep the system safe. We don't have to build something completely new.

Sergiy Sumnikov
HOST

I can add from our experience that black-box outputs with poor explainability are often an adoption killer too. If a system outputs a risk score of, say, seventy-eight percent with no context, clinicians can't really trust it, because in the end it's the clinician, not the model, who is responsible for the outcome, and who has to justify it to colleagues. So it helps to provide credibility signals. Doctors are trained skeptics; they treat everything with a healthy dose of skepticism. And in some high-risk settings it's even worth adding friction before the next decision step.

Let's continue. What role do interface design and workflow integration play in preventing misuse or over-reliance?

Dmytro Loposhansky
GUEST

It's really important to integrate well with the workflow first. If you build something completely outside the actual clinical workflow, no one will use it. But even when you are integrated, the harder question is whether you're preventing misuse or over-reliance, and that's genuinely hard to nail. When a tool works really well, people tend to trust it and stop questioning it. We see that everywhere, not just in healthcare: good AI tools make life easier precisely because you no longer spend as much effort on the task.

So there's a balance. Do clinicians still need to reason it through themselves and verify the model from time to time? There are a few ways to approach this. One is to design the workflow so the clinician makes their decision first, and only then does the tool offer its judgment, so the two can be compared. Another is not having the tool do a hundred percent of the work. Take tools like OpenEvidence or UpToDate, which are mainstream among clinicians. They're question-and-answer systems, and the final decision still sits with the clinician. The system doesn't hand down a diagnosis; it surfaces the helpful resources the person would otherwise spend a long time digging up.

That's probably the sweet spot, not fully autonomous, but still leaving the person to make the final call. Of course, if your tool works so well and is so thoroughly tested that no one needs to doubt it anymore, maybe that's a different path. Who knows.

Sergiy Sumnikov
HOST

Let's go a bit deeper. How much of trust comes from the model itself, versus how the system is designed and presented?

Dmytro Loposhansky
GUEST

Let me bring a specific example. We built a model, put it into production as a separate application with its own set of credentials, and then hoped people would use it. In hindsight there was no way that would succeed. Nobody can navigate a hundred separate systems, each with its own login. People care about the result and the outcome, and clinicians treating patients care about that above all. They don't want to juggle credentials and log into different systems.

So the goal is to simplify integration as much as possible. We integrated the model into the Epic Hyperspace environment and turned the tool into a button you click. We passed a deep link, so the user is already authorized in Epic and there's no second login into our app. That simplified everything, and it shows how much the design aspect drives actual adoption.

Sergiy Sumnikov
HOST

Let's dig into that. What are common design mistakes that make clinicians reject otherwise useful systems? Maybe one or two examples.

Dmytro Loposhansky
GUEST

The way the clinical environment is built today pushes a lot of data onto clinicians and asks them to make the judgment. Picture being on the floor with ten screens and an enormous amount of input, and you

have to decide. A lot of tools assume the answer is to provide as much information as possible to the person deciding. Unfortunately that's not how it works. Clinicians are already overwhelmed, and the mental load is enormous. That's one of the key design flaws: pushing too much complexity all at once.

The complexity matters, but it has to be layered. Give a more abstract view first, one you can read instantly to understand the state of the patient, and then let people drill into specific areas when they need to, because being specific still matters. The judgment shouldn't rest on surface information alone, but it does need those layers.

In general, a lot of healthcare systems date back to the seventies and eighties. The manufacturers built them, got regulatory approval once, and never updated them, because getting approval again is expensive and hard, so why change something people already use? There's a real aversion to doing things differently for its own sake, which means you need a clear benefit over the current state. Even Epic's UI has an enormous amount on screen. One of the projects we're working on is an overlay on top of Epic that serves clinician judgment better, because Epic is used by administrators, billing, and clinicians across every department. It's sophisticated and heavy, and it doesn't always match what a given person is trying to do.

Sergiy Sumnikov
HOST

As a UX design agency, I'll add a few adoption killers we see with our customers, whether they're innovators inside hospitals or health-tech startups building AI for hospitals.

In some cases products are designed for demos, not for real environments. And real environments are noisy. In the ICU there are interruptions, people are wearing gloves, the lighting is bad, and we've already touched on alarm fatigue. Keeping those conditions in mind is essential for a successful deployment. Two smaller points: we've seen AI positioned as overconfident, and sometimes underconfident, and neither is right, so it's better to display confidence levels and uncertainty boundaries honestly. And one piece of feedback from innovators at a hospital was that current value alone isn't enough; they need temporal context, because medicine trends over time, and a rising risk is often more important than an absolute value.

I have a few more questions on trust and safety, but we're out of time for this section, so let's move to ROI, which is more organizational.

04 Return on Investment

Who actually decides on deployment, which metrics matter to leadership, and how to prove value before you're fully deployed.

Sergiy Sumnikov
HOST

Let's start with who actually decides whether an AI solution gets deployed in a hospital.

Dmytro Loposhansky
GUEST

It varies a lot by the size of the system, and there are many structures to map out, but the north star is this: if you're deploying an AI tool, IT is almost always involved, plus the executive responsible for the department you're deploying into. For a clinical AI tool, that's IT and a clinical VP or the chief medical information officer. For an HR tool, it's IT plus the VP of Human Resources, and so on.

It's not only about the VPs, though. You can sell to individual departments, individual teams, even individuals. Everyone has their own needs; it depends on the scale you're targeting. The VP level tends to come in when the conversation is about deploying wall-to-wall for everyone. When the system is only needed within a specific department or workflow, the head of that service line or department can often make the decision.

It's fairly similar across large organizations: each team has a budget allocated centrally every year, and you work with that. Depending on how well your tool scales, you go and talk to progressively more senior managers. At smaller institutions it's different: one person tends to be responsible for more of the operation, so you can talk to fewer people and still make the case.

Sergiy Sumnikov
HOST

If you had to justify your AI product to leadership, what metrics would matter most?

Dmytro Loposhansky
GUEST

I don't think there's a single one. As I said, there's no one-size-fits-all here. Time saved, maybe. If you can show you're saving a meaningful amount of time for a very expensive specialist, that's a great metric. But I can name plenty of cases where time saved wasn't the number that made the difference.

Ambient AI is a good example. People initially assumed it would save clinicians a lot of time, but even during the pilot stage you could see there weren't huge time savings, because the doctor still has to review and sometimes correct the note. The real value is that they no

longer have to focus on note-taking during the visit. That raises the quality of care and makes clinicians feel less burdened, so the value shows up in a different layer.

That's why you have to be customer-obsessed: focus intensely on who you're trying to help and what they actually need. The right metric is the one that matters most to the person you want to help.

Q&A

Questions & Close

Proving value before full deployment, and a note on where to take the conversation next.

Sergiy Sumnikov
HOST

We're done with the main part, so let's open the Q&A. A few quick actions first, which I've posted in the chat: if you'd like custom feedback from Dmytro, forward your questions to my email and they'll go into the follow-up. If you'd like help improving the adoption and outcomes of your AI products, get in touch with us as a specialized UX agency. I've also posted Dmytro's LinkedIn profile, which is worth following, and my own for upcoming webinars.

While people think of questions, I have one more on my own list, on ROI. How can teams demonstrate value before full deployment, and what does a compelling ROI story actually look like in practice?

Dmytro Loposhansky
GUEST

It comes down to evidence. That's not groundbreaking, but if you want to show your tool works, then show it. Use case studies. If you've already done these integrations at other sites, point to those results; that speaks louder than any marketing video. Of course there's a cold-start problem: how do you get the first one when everyone expects five case studies? You show, in principle, how the problem can be solved. If the problem is painful enough, health systems can see it, and you only need to find one that understands and prove technically that you can deliver.

The best ROI story is simple: who had the problem, what changed, and what it's worth. There's a Cleveland Clinic case study published recently on optimizing incoming referrals. If you work in a health system, you know referrals come in by fax, and coordinators review them, match them to the right patient chart, and update everything by hand. The system they adopted automated that. Their pitch was very simple: after deployment we automate eighty percent of that workflow, which frees up three full-time coordinators, and here's

what that costs. If you can tell a story that simple and that compelling, that's really all there is to it.

The best ROI story is simple: who had the problem, what changed, and what it's worth.

DMYTRO LOPOSHANSKY

Sergiy Sumnikov
HOST

Wonderful. I don't see any further questions, so I'd like to thank you, Dmytro, for taking the time and accepting my invitation, and to wish everyone a great day or evening, wherever you are.

Dmytro Loposhansky
GUEST

Thank you very much, Sergiy, and thanks to everyone for joining. I hope it was useful, and I'd be happy to stay in touch.

Sergiy Sumnikov
HOST

Take care. Bye now.

End of transcript. Recorded 15 April 2026 · Halo Lab · halo-lab.com