SG AI Solutions

A FOUNDER'S FIELD GUIDE

The 7 Questions to Ask Before Signing a Healthtech AI POC

A 60-second self-scoring test, then the engineering depth behind each answer.

Saber Graf, PharmD, PhD Neurosciences
Healthtech AI architect & engineer

Why these seven

> **Most healthtech AI POCs don't die on the model. They die on a question nobody asked before signing.**

The demo is impressive. Vendor proposals start to look interchangeable. The POC stalls before production, and the runway burns on a thing that was never going to ship.

These are the seven questions I ask before I let a client sign anything. None of them require you to be technical. They require the other side, a vendor or your own team, to be specific. Score yourself in sixty seconds, then read the depth behind each question.

I'm a pharmacist by training, a neuroscientist by PhD, and an engineer by daily practice. *These are the questions that combination taught me to ask before money moves.*

The 60-second test

Tick YES or NO. Each question maps to one deep-dive in the pages that follow.

1. **The problem is worth AI.**
   YES if you can name the decision or task AI will change, and what a non-AI baseline would cost. Y / N

2. **The data exists.**
   YES if the data is already collected, you know its volume and quality, and you have the right to use it. Y / N

3. **Regulatory ownership is named.**
   YES if one person owns the regulatory and clinical responsibility, on paper, today. Y / N

4. **The real cost is on the table.**
   YES if you have a number for production, maintenance, and retraining, not just the POC. Y / N

5. **Ownership is unambiguous.**
   YES if you know who will own the model, the data, and the production stack after the POC. Y / N

6. **Success is defined clinically.**
   YES if you have a success metric that means something to a patient or clinician, not only an F1 score. Y / N

7. **There is an exit plan.**
   YES if you know what happens, technically and contractually, if it doesn't work. Y / N

6–7 → Go
Signing with your eyes open. Use the deep-dives to sharpen the contract.

3–5 → Rework
The idea may be sound, the setup isn't. Fix the NOs before money moves.

0–2 → No-go
Signing today buys a demo and a sunk cost. See what to put in place first.

Every NO above is a section below, with the red flags to watch and the answers a serious vendor gives.

QUESTION 1

1 Is the problem worth an AI solution?

> Before "can we build it", ask "should anything be built at all". A surprising share of healthtech AI POCs solve a problem a spreadsheet, a rule, or a single clinician already solved.

WHAT YOU'RE REALLY ASKING

Are you buying a capability, or buying a narrative? AI earns its place when the pattern is real, repeated often enough to matter, and expensive to handle by hand. If the task happens twice a month and a junior analyst clears it in an afternoon, an AI budget is the wrong tool, however good the demo looks.

THE TYPICAL TRAP

The problem gets reverse-engineered from the technology. Someone has access to an LLM and goes looking for somewhere to point it, instead of starting from the decision that needs to change. In healthtech this is expensive, because the validation and data work around even a small model is never small.

RED FLAGS

- No non-AI baseline has ever been computed. Nobody knows what the manual or rules-based version costs.
- The ROI story only works at a scale you don't have yet.
- The internal champion can't state, in one sentence, the decision the model will change.
- "AI" is in the pitch title but absent from the actual workflow.

GOOD ANSWERS TO EXPECT

- A named decision or task, with its current cost in time, money, or error rate.
- A baseline (manual, rules, or simpler statistics) the AI is explicitly expected to beat.
- A clear "if the model gets this right, X changes for Y people".

FROM THE FIELD

A team wanted an LLM to triage a weekly intake of around two hundred rows. A sorting rule handled ninety percent of it in seconds. The AI budget was about to be spent on the remaining ten percent, which turned out not to need a model at all. The right answer was a rule plus one human review, not a POC.

QUESTION 2

2 Does the data actually exist?

> The most expensive sentence in healthtech AI is "we'll get the data later". This is where POCs quietly die, months after the contract is signed and the budget is committed.

WHAT YOU'RE REALLY ASKING

Do you have a dataset, or a hope? Three axes decide it: volume, quality, and the legal right to use the data for this specific purpose. All three, not one out of three.

THE TYPICAL TRAP

The data exists, technically, but is unusable in practice. Wrong format, no labels, trapped in a partner's system you can't query, or collected under a consent that doesn't cover model training.

RED FLAGS

- "The data is somewhere in the EHR" or "in the LIMS", with no one who has actually pulled it.
- No labeling plan and no labeling cost, when the task needs labels.
- Nobody has counted the actual usable rows.
- The legal basis (consent scope, processing purpose, agreements with any processor) is assumed rather than checked.

GOOD ANSWERS TO EXPECT

- A real count of usable, task-relevant samples, not total records.
- A labeling strategy, and a clear answer on who pays for it.
- A documented legal basis for using the data for AI: GDPR purpose, consent scope, a data processing agreement with any third party.
- Honesty about class imbalance and edge cases.

FROM THE FIELD

The gap above is real. I have also seen the opposite problem from the inside: in my own published work on physiological-signal classification, a meaningful share of the training labels were simply wrong, and fixing that label noise mattered more than any change to the model (Scientific Reports, 2025).

QUESTION 3

Who owns regulatory responsibility?

> In healthtech, "the AI decided" is not a defense. Someone owns the outcome. Make sure that someone is named before you sign, not discovered after something goes wrong.

WHAT YOU'RE REALLY ASKING

Is this tool informational, decision support, or a medical device? The answer changes who is liable and how high the bar sits. This is the question where you involve your regulatory or quality lead, or a lawyer. This document is not legal advice, and the point here is only to make sure the question is owned.

THE TYPICAL TRAP

Regulatory status is treated as a "later" problem. Then the intended use of the POC, the thing it is actually meant to do, turns out to already place it in medical-device territory, with documentation and validation nobody budgeted for.

RED FLAGS

- Nobody can say whether the intended use makes this a medical device.
- "It's just decision support", with no line actually drawn between support and decision.
- No named human owner for clinical and regulatory responsibility.
- A vendor who offers to "handle compliance" but won't put scope in writing.

GOOD ANSWERS TO EXPECT

- A named human who owns regulatory and clinical responsibility.
- An explicit intended-use statement.
- Awareness of the relevant frame, whether SaMD, MDR, IVDR, or the AI Act risk class, even if a specialist finalizes it.
- A plan to bring quality and regulatory in early, not at the end.

FROM THE FIELD

A "wellness dashboard" was scoped as a simple analytics view. Its intended use, in the actual pitch, was to flag patients who needed follow-up. That quiet phrase moved it toward medical-device territory the team had not planned or priced for. The fix was cheap when caught early, and would have been very expensive caught late.

QUESTION 4

What is the real total cost?

> The POC price is the cheapest number you will ever see on this project. Ask for the other three, production, maintenance, and retraining, before you fall in love with the demo.

WHAT YOU'RE REALLY ASKING

Are you funding an experiment, or a system? In healthtech especially, the cost of running, monitoring, and retraining a model in production usually dwarfs the cost of the POC that proved it could work once.

THE TYPICAL TRAP

The POC is priced, approved, and delivered. Everything after it, hosting, drift monitoring, retraining, on-call, is "to be discussed", which in practice means unbudgeted. The project then stalls at exactly the moment it was supposed to create value.

RED FLAGS

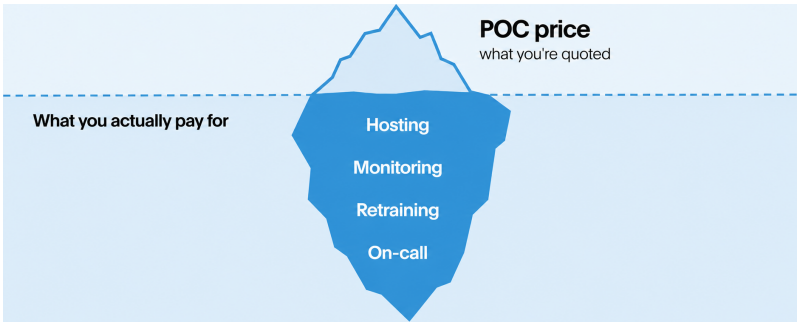
- A clean price for the POC and silence on everything that comes after it.
- No line item for monitoring, infrastructure, or retraining.
- Cloud and inference costs not modeled at real, production-level volume.
- "Once it's trained, it just runs." It does not, and in regulated healthtech least of all.

GOOD ANSWERS TO EXPECT

- A cost view across the full path: POC, then production, then maintenance, then retraining.
- Named infrastructure and inference costs at the volume you actually expect.
- A retraining trigger and cadence, and who pays for it.
- A clear answer to "who carries the cost when the model degrades".

FROM THE FIELD

A fifteen-thousand-euro POC was approved without friction. The production reality, hosting, monitoring, quarterly retraining, and a person on call, came to a multiple of that POC every year. None of it was hidden maliciously. It simply never came up, because no one asked the question on this page.



QUESTION 5

Who owns the model, the data, and the stack?

> Three different ownerships, three different traps. Lock-in hides in the gaps between them, in the clauses you didn't think to read.

WHAT YOU'RE REALLY ASKING

After the POC, what do you actually hold? The model weights and artifacts? The training data and the data derived from it? The pipelines that run it? And, underneath all of that, the right to walk away and take your work with you?

THE TYPICAL TRAP

The vendor owns the model and the pipeline, you own the risk and the bill, and the cost of switching is quietly engineered to be high. Every improvement you fund makes leaving more expensive.

RED FLAGS

- "Our proprietary model", with no clause on what you keep.
- Your data improving a model you do not control, with no say in it.
- No export path for weights, artifacts, or derived datasets.
- A production stack that is a black box only the vendor can run.

GOOD ANSWERS TO EXPECT

- Explicit ownership of model artifacts, training data, and derived data.
- Portability and exit clauses written into the contract.
- A clear statement on whether your data trains a shared model used by others.
- A stack that you, or a competent third party, could operate if you had to.

FROM THE FIELD

In one contract, every label the client paid its own staff to create silently became the vendor's asset. The client had funded the most valuable part of the system, the labeled data, and ended up renting access to it. That single clause set the price of ever leaving.

QUESTION 6

What does success look like, clinically?

> A model that is ninety-five percent accurate can be clinically useless. Define the metric that means something to a patient before you celebrate the one that looks good on a slide.

WHAT YOU'RE REALLY ASKING

Will you measure statistical performance, or real-world utility? They are not the same number, and the gap between them is where healthtech AI projects lose credibility with the clinicians who have to use them.

THE TYPICAL TRAP

Success is defined as an offline score, accuracy, AUC, F1, on a clean held-out test set. That score is disconnected from the clinical decision the tool supports and from the very different costs of a false positive versus a false negative.

RED FLAGS

- One headline accuracy number, no confidence interval, no breakdown by subgroup.
- No distinction between false positives and false negatives where that distinction is the whole point.
- Success defined by the vendor's metric rather than by the clinical workflow.
- No plan to evaluate prospectively, in real conditions, only on a static test set.

GOOD ANSWERS TO EXPECT

- A metric tied to a clinical decision and to the asymmetric cost of each error type.
- Performance reported by subgroup, so a silent bias can't hide in the average.
- A plan to evaluate in real conditions, not just on a held-out split.
- A "good enough to deploy" threshold agreed up front, not negotiated after the fact.

FROM THE FIELD

A model scored beautifully overall and failed precisely on the rare, high-stakes cases it had been bought to catch. The average was excellent. The clinical value was close to zero, because the average was carried by the easy cases a clinician never worried about in the first place.

7

QUESTION 7

What's the exit plan if it doesn't work?

> This is the serious question that separates a partner from a vendor. Decide the off-ramp while everyone is still optimistic, because no one will want to design it once things are going badly.

WHAT YOU'RE REALLY ASKING

If the POC fails, or the partnership ends, what do you keep, what do you owe, and how fast can you stop? A project without an exit plan doesn't end, it lingers.

THE TYPICAL TRAP

No kill criteria. Because no one defined what failure looks like, a failing POC drags on. It gets renewed "to give it a chance", and a year of runway disappears into a thing everyone already suspected wouldn't work.

RED FLAGS

- No pre-agreed kill criteria and no decision date.
- No off-boarding clause for data return and artifact handover.
- Data and artifacts you can't actually retrieve on exit.
- Auto-renewal with no real review gate.
- "Let's just see how it goes."

GOOD ANSWERS TO EXPECT

- Explicit go, rework, or no-go criteria, with a decision date attached.
- An off-boarding clause: data returned, artifacts handed over, transition support defined.
- A defined notice period.
- Shared agreement that "no-go" is an acceptable, planned outcome, not a failure of the relationship.

FROM THE FIELD

Two teams, the same failing POC. One had written kill criteria and a decision date, reviewed honestly, and walked away inside a month with its data in hand. The other had nothing written, renewed twice, and lost the better part of a year before reaching the same conclusion the criteria would have reached on schedule.

Where this leaves you

Look back at the seven questions. Not one of them is about the model. Every one is about whether the people around the model, the vendor, your team, you, have been specific.

A proposal that answers all seven crisply is rare, and worth signing with. One that dodges them is selling you a demo. And the good news is that you don't need to be technical to ask any of this. You need the other side to be precise, and these questions force precision.

So if you scored a three out of seven today, that is not a verdict. It is a punch list. Fix the NOs, and you turn a risky signature into a deliberate one.

Before you sign

If you have a real AI project to validate, ship, or de-risk in the next few months, the fastest way to put these seven questions to work on your specific case is a short call.

Book a 30-minute call
calendly.com/saber-graf-sg-ai-solutions/30min

Not ready for a call? Every red flag here becomes a short post, follow along on LinkedIn.
linkedin.com/in/saber-graf

Read the full reasoning, with the engineering underneath, in the book in progress
book.sg-ai-solutions.com

Saber Graf (PharmD, PhD Neurosciences) · Healthtech AI architect & engineer, SG AI Solutions · sg-ai-solutions.com