

What We've Learned Across Our Governance Sprints

Anonymized aggregate from health systems that ran a structured AI governance sprint with us.

Where the time goes. What's always missing. What the fast ones do differently.

No system named. No number reported at an n small enough to identify anyone.

METHOD

What a sprint is, and what this isn't

- A fixed-window pass over a **real intake queue**
Real tools, real vendors, real reviewers, real deadlines. Not a maturity assessment and not a survey.
 - The same structured evidence request every time
Which is what makes the submissions comparable to each other, and what makes a pattern across systems mean anything.
 - Observations, not research
There is no control group and there was never going to be one.
-

Self-selection is the honest limitation. Every system in this set chose to run a sprint, which biases the whole thing toward organizations already taking governance seriously. Read these as a floor.

Read the ordering, not the decimals.

The proportions move sprint to sprint. The **rank order has not moved once** — at systems of very different sizes, academic and community, with and without a Chief AI Officer.

So if you're holding your own program up against this, compare shapes. If your bottleneck sits somewhere other than where it sits on the next slide, that difference is the finding worth chasing.

WHAT THIS IS NOT

Not industry data. This is our own book, anonymized, from a self-selected sample. Anyone quoting it as a market benchmark is overreading it, including us.

Where the calendar actually goes

STAGE, IN ORDER OF ELAPSED TIME	WHO IS WORKING ON IT
Waiting on vendor evidence	Nobody
Waiting for the next meeting	Nobody
Chasing the internal sponsor for a usable use case	One person
Reviewers reading and writing	The committee
Discussion in the room	The committee
Decision, record, and communicating it	Chair / staff

Rows one to three are queue. Rows four to six are work. The ratio between them is the one number I would put on a governance dashboard, and almost nobody measures it.

Most of the delay happens while nobody is working.

In every sprint we have run, most of the elapsed time on a typical item was spent with **nobody working on it**. It sat waiting for a vendor to send a document, or waiting for a meeting that happens on a fixed calendar whether or not anything is ready to discuss.

That matters because the standard response to a slow committee is to ask the committee to work harder — more meetings, more reviewers, faster reads. That addresses the smaller half of the elapsed time and leaves the rest untouched.

THE REFRAME

Measure time-in-queue separately from time-in-review. The first is a scheduling and intake problem you can fix inside a month. The second is a capacity problem that costs money.

What is missing, ranked

DOCUMENT, MOST OFTEN MISSING FIRST	USUALLY BECAUSE
Post-deployment monitoring plan	Nobody has decided who owns it after go-live
Subgroup performance breakdown	It was never run
Described failure modes	Answered as an uptime and escalation question
Model update and change-notification terms	It's a contract term nobody asked for
Consent and disclosure workflow	The vendor treats it as the customer's problem
Data retention and training use	Answered with a policy link, not the contract

This order has been stable across every sprint. If your own list ranks differently, that is worth a conversation — I would like to know about it.

The monitoring plan is missing for a structural reason

It is the only document on that list that a vendor **cannot produce alone**. It requires somebody inside your organization to commit to measuring something forever.

- A vendor can supply metrics. They cannot supply an owner.
And the owner is the part that makes it a plan rather than a page.
- That owner usually has no budget line for it.
Which is why the plan exists at approval and has evaporated by day 90.
- Nothing in most governance processes checks whether it ever ran.
Approval gets audited. Monitoring gets assumed.

What the fastest systems do differently

- 1 They publish the evidence request before a vendor is selected.

Moves the waiting to before the clock starts.

- 2 They let staff return incomplete submissions without a meeting.

The single largest queue reduction we have measured.

- 3 They tier at intake, so most items never reach the full committee.

Most intake at most systems never needed the full room.

- 4 They write the decision record in the meeting, not after it.

A record written a week later is a summary of a memory.

- 5 They hand the sponsor a template, not a form.

Sponsor-side delay is a writing problem, not a willingness problem.

What did not predict a faster or better program

VARIABLE	THE ASSUMPTION	WHAT WE SAW
Bed count	Bigger systems are further along	No relationship in either direction
Having a Chief AI Officer	The title unblocks the program	Helps escalation. Does not move throughput.
Which framework they adopted	NIST vs. accreditor vs. homegrown decides quality	Close to invisible in our set
Vendor size	Large vendors submit better evidence	No. Having been asked before predicts it.
Number of committee members	More reviewers, more throughput	Inverted past a certain size in our set

The variable that did track: whether one named person owned the queue between meetings. Not their title. Whether the job existed at all.

THREE WAYS A PROGRAM STALLS

You will recognize at least one of these

THE POLITE QUEUE

Nothing is ever refused. Incomplete submissions are accepted and “followed up on.”

Throughput looks healthy until you count how many approvals were conditional on a document that never arrived.

THE HEROIC CHAIR

One person reads everything, writes everything, chases everything.

It works, visibly, for about nine months. Then that person changes jobs and the program has no memory.

THE INVENTORY THAT ISN'T

A spreadsheet of tools. One row each. No use cases.

It answers “what did we buy.” It cannot answer “where are we exposed,” which is the only question anyone asks during an incident.

What we got wrong.

I thought the hard part was judgment. In most sprints it isn't. The committees I've watched make good calls when the material is in front of them. The failure is that the material arrives late, incomplete, and in a shape nobody can compare against last month's.

I thought a sandbox was the gap. For a small number of tools it genuinely is. For most of what actually comes through intake, the risk lives in the deployment rather than the model — and evidence plus monitoring buys more safety per dollar than a testing environment does.

STILL UNRESOLVED

I have no good answer for the systems doing the hard asking. They produce evidence the whole market then free-rides on, and nothing compensates them for it.

Five things that cost nothing

- 1 Measure time-in-queue separately from time-in-review.
One month of data tells you which problem you have.
 - 2 Publish your evidence request on your own website.
Vendors prepare once. You stop negotiating the list per submission.
 - 3 Authorize staff to return incomplete submissions.
No meeting, no debate, standing rule, in writing.
 - 4 Add one line to every approval: who owns monitoring, first check date.
Then actually check on that date. That's the whole control.
 - 5 Count how many items on your last agenda needed the full room.
That number is your tiering business case, already written.
-

Nobody can see across health systems from inside one.

That is the only reason this document exists. Every system in it has a clear view of its own program and no view of anyone else's, which makes it genuinely hard to know whether your bottleneck is unusual or universal. Enough sprints in, most of them are universal.

Troy Bannister — Founder & CEO, Onboard AI

Anonymized aggregate, August 2026. No system named. If your numbers look different from these, I would rather hear it than not.