HOSPITAL AI

Measuring AI when nothing clinical is claimed

Baseline before launch, a fair comparison, hand-sampled quality, adoption, full cost and a stop rule agreed in advance.

Swipe →

Gaurav Phogat
gauravphogat.com



THE PROBLEM

Approved on a slide, judged on anecdote

- Non-clinical AI needs no ethics committee, so it launches on enthusiasm
- Six months later, finance asks whether it worked
- The team shows usage charts and happy quotes
- Nobody can say what changed, compared with what, at what cost

Gaurav Phogat
gauravphogat.com



WHY MEASURABLE

Operational work is easier to measure, not harder

- How long enquiries wait
- How many claims return for missing documents
- Hours a coordinator spends on a report
- Draft articles needing major rework

The hard part is discipline: decide before launch, collect a baseline, do not move the measure.

Gaurav Phogat
gauravphogat.com



DECISION FIRST

Start with the decision, not the dashboard

DECISION	Expand, keep as is, stop, or redesign and retest
CFO	Net effect on cost or revenue with a credible comparison
UNIT HEAD	Did workload and service actually improve
OPERATIONS	Did quality hold up
STOP RULE	Written in advance, or every result becomes promising

Gaurav Phogat
gauravphogat.com



FOUR FAMILIES

Have at least one measure from each

1. Outcome: time to first response, claim rejections, enquiry conversion
2. Quality: sampled and reviewed by people, not model confidence
3. Adoption: do intended users keep using it after the novelty
4. Cost to run: licences, integration, review, supervision, internal team

Outcome without quality is dangerous. Outcome without cost looks profitable in pilot, expensive at scale.

Gaurav Phogat
gauravphogat.com



BASELINE

A result only means something against a comparison

- Measure for several weeks before go-live, same definitions
- Compare against a unit, shift or work share on the old process
  Hospitals have seasons; a quiet-month launch looks brilliant
- Write each definition down and hold it still

Gaurav Phogat
gauravphogat.com



QUALITY

Why AI quality needs human sampling

- Output can look excellent and be wrong
- Define good enough in plain terms per task
- A small group scores a random sample weekly with protected time
- Watch the trend: quality drops on new inputs or vendor model updates

Gaurav Phogat
gauravphogat.com



HIDDEN COSTS

Adoption and the review load

- Falling adoption is an early signal: slow, needs fixing, does not fit
- Time saved drafting can reappear as supervisor review time
- Early numbers mislead: enthusiasts, low volume, light review
- Measure under realistic volume, not in a friendly corner

Gaurav Phogat
gauravphogat.com



REPORT HONESTLY

09

"

Report only what was measured, in the words it was measured in. Never imply a health benefit from a project evaluated as an operational one.

Faster discharge paperwork is not better patient outcomes. Report the failures too.

Gaurav Phogat
gauravphogat.com



READ THE FULL GUIDE

Measuring non-clinical AI projects, in full

- The one-page measurement plan to sign before kick-off
- Who samples quality, and how often
- Reporting upward without overclaiming
- gauravphogat.com/measuring-non-clinical-ai-projects

Gaurav Phogat
gauravphogat.com

