

The Jonesi Consultancy | Houston, TX

LEADERSHIP RESOURCE · MEASUREMENT GUIDE

AI ROI: QUESTIONS BEFORE YOU MEASURE



Defining Outcomes Before Calculating Returns

Measurement is not a reporting exercise. It is a decision you make up front.

Most AI return is undefendable for one reason: nobody wrote down what would change before anything changed. Capture the baseline first.

WHY AI RETURN IS HARD TO DEFEND

Ask a leadership team what their AI investment returned and you will usually get one of two answers. Either a confident number that dissolves under the first follow-up question, or an honest admission that nobody can say. Both come from the same place. Success was defined after the spending, which means it was defined by whatever happened to be available.

This is a discipline problem rather than a mathematical one. The arithmetic of AI return is not difficult. What is difficult is committing to a measure before you know whether it will flatter you.

1 Why the Number Goes Vague

No Baseline Was Captured.

Somebody is certain the process is faster and nobody wrote down what it cost before. Without the before, the after is an anecdote with a decimal point.

The Measure Was Invented For The Initiative.

A metric nobody tracked before the project will not be tracked after it. Within two quarters it quietly disappears, usually around the time it stops being flattering.

Time Saved Was Self-Reported.

People asked how much time a tool saved them consistently overestimate, not dishonestly but because the comparison is against a remembered version of the old process.

The Gain Never Left The Individual.

An hour saved that gets absorbed into a busier day is real for that person and invisible in financial statements. This is the most common gap between what staff experience and what leadership can find.

Attribution Was Never Settled.

The process also got a new manager, a cleaner handoff, and a template. Claiming all of the improvement for AI is how a credible number becomes a disputed one.

THE GOVERNING STANDARD

Defensible logic beats false precision. A number you can walk a skeptical CFO through, with its assumptions visible and its limits acknowledged, survives scrutiny. A more impressive number built on estimates nobody can trace does not, and it costs you credibility on the next request.

2 7 Questions to Answer First

Answer these before implementation, not after. Every one of them becomes materially harder to answer honestly once money has been spent.

1

Which specific workflow do we expect to change?

Name a process with a start, an end, and an owner. "Our team will be more productive" cannot be measured because it is not a process.

WHY IT MATTERS: ROI needs a unit of work. "Use AI more" is not an operational outcome.

- › What task, process, handoff, or decision should change?
- › Who performs it today, and how often?

2

What does that workflow cost us today?

Hours, headcount, error rate, rework, cycle time, or delay. Capture it now. This is the single step most organizations skip and the one that cannot be recovered later.

WHY IT MATTERS: Without a baseline, improvement cannot be separated from normal variation.

- › How much time, rework, delay, external spend, or risk exists today?
- › What evidence already exists to document the current state?

3

How do we know that number is real?

Pulled from a system, sampled over two weeks, or estimated by the person who does the work? Write down which. The strength of the whole calculation is the strength of this figure.

WHY IT MATTERS: Useful measurement should survive scrutiny beyond the project team.

- › What observable change would be difficult to dismiss as enthusiasm or self-report?
- › Which measures can be captured consistently before and after implementation?

4

What improvement would make this worth doing?

A threshold set in advance, before you know the result. If any improvement counts as success, you are not measuring, you are justifying.

WHY IT MATTERS: A useful target describes a business change, not merely an AI activity.

- › Should cycle time fall, quality improve, capacity increase, or risk decrease?
- › What would leadership consider meaningfully better?

5

Why does that improvement matter to the business?

Trace it to capacity, cost, revenue, risk, or quality. An improvement that connects to none of those may still be worth having, but it is not a return.

WHY IT MATTERS: Not every efficiency gain is strategically valuable.

- › Which objective, service standard, cost pressure, or strategic priority does this support?
- › What happens if the workflow does not improve?

6

Who owns the measurement, by name?

Not a department. A person, with the access to pull the number and the standing to report it unflatteringly.

WHY IT MATTERS: Metrics without ownership usually disappear when the rollout becomes busy.

- › Who will collect, validate, and review the evidence?
- › Who has authority to act when the results are weaker than expected?

7

Over what period, and when do we look?

Set the review date now. Ninety days is usually long enough for a habit to form and short enough that people still remember the baseline conversation.

WHY IT MATTERS: Immediate enthusiasm and sustainable operational impact are not the same thing.

- › What can reasonably be observed in the first 30 days?
- › What needs 60 to 90 days, or longer, to stabilize?

Before the First ROI Calculation



Outcome Defined

We can name the workflow or decision expected to change.



Baseline Available

We have credible evidence of the current state.



Business Relevance Established

We know why the change matters beyond AI adoption itself.



Measure Selected

We know what evidence will demonstrate improvement.



Owner Assigned

Someone is responsible for collecting and reviewing the evidence.



Review Window Set

We know when early and sustained results will be assessed.

3

Activity Is Not Outcome

The most common measurement mistake is reporting evidence that something happened as evidence that something improved. Activity metrics are easy to collect, and that is precisely why they dominate the first year of most AI programs.

ACTIVITY METRICS	OUTCOME METRICS
 Licenses deployed	 Cycle time on a named process
 Users logged in this month	 Rework and correction rate
 Prompts submitted	 Output quality, sampled and scored consistently
 Training sessions delivered	 Capacity released, and where it went
 Attendance and completion rates	 Errors caught before they reached a client
 Self-reported hours saved	 Time from request to delivery, measured in a system
 Tools evaluated	 Decisions made faster with the same or better quality

Activity metrics are not worthless. They tell you whether anything is happening, which is worth knowing early. They simply cannot answer the question a board asks, which is what changed in the business. Report them as leading indicators and label them as such.

THE CAPACITY QUESTION NOBODY ASKS

If a process now takes four hours instead of six, where did the two hours go? Into higher volume, into work that was previously deferred, into reduced overtime, or into a slightly easier week? All four are legitimate. Only some of them are a financial return, and the difference matters enormously to whoever has to defend the investment.

A USEFUL TEST

If the metric can increase while the underlying work remains unchanged, it is probably an activity measure rather than an ROI outcome.

4

Write It Down Before You Spend

One row per workflow. Complete it before implementation and give a copy to whoever approves the expenditure. This single page prevents most of the arguments that happen at the end of the year.

WORKFLOW OR DECISION	BASELINE TODAY	TARGET AND BY WHEN	OWNER

3 Rules for the Table Above

Baseline First.

If you cannot fill the baseline column, you are not ready to spend on that workflow yet. Go and measure it first. It usually takes two weeks.

Use What You Already Track.

Use a measure that already exists in a system somebody already looks at. Inventing one guarantees it will be abandoned.

Name A Person.

The owner is a person, not a function, and they should be comfortable reporting a result that does not flatter the decision.

ON HONESTY IN THE TARGET COLUMN

A target set after you have seen the result is not a target, it is a description. The value of writing the number down in advance is entirely that it can turn out to be wrong.

These seven questions will carry an organization a long way, particularly for the first few workflows. There is a point at which they stop being sufficient.

› You are measuring across several departments and need results that can be compared rather than each defended separately.

› The gains show up in quality, risk, and decision speed rather than in hours, which are harder to put in a number a CFO will accept.

› You need to report to a board, a funder, an owner, or an audit committee whose job is to find the weakness in the numbers.

› Multiple things changed at once and attribution has become genuinely contested.

› Productivity, quality, adoption, workforce, risk, and financial effects have to be separated without double-counting them.

› You are being asked to forecast return on an investment that has not happened yet.

OUR 7-LEVEL ROI FRAMEWORK

It distinguishes between levels of evidence, from time saved on a task through to financial impact that appears in the statements, so that a claim can be made at the level the evidence actually supports rather than at the level the conversation would prefer. That distinction is what makes an ROI number survive scrutiny.

THE HABIT THAT MATTERS MORE THAN THE FRAMEWORK

Capture the baseline. Organizations with a mediocre measurement approach and good baselines can always improve their analysis later. Organizations with a sophisticated framework and no baselines have nothing to analyze, and no amount of subsequent rigor recovers what was not recorded at the time.

IF YOU TAKE ONE THING FROM THIS DOCUMENT

Take two weeks and measure what your three most important processes cost you today. Do it before the next tool decision. Everything downstream gets easier, and it is the only step in the whole exercise that cannot be done retroactively.

© 2026 The Jonesi Consultancy. All rights reserved.

Building a measurement case your CFO will sign off on.

Our approach is built by a CPA, which is why it survives the follow-up questions.

