Buying an AI Tool *Playbook.*

A decision protocol for adding an AI tool to an operating business.

Built for one decision meeting, durable across tool cycles. Made to be printed, marked up, and used in the room where the call is made.

CONTENTS

Contents

THE FRAME

THE FOUR GATES

THE RECORD

THE FRAME · 00

Start here

A decision protocol for adding an AI tool to an operating business. Built for one decision meeting. Durable across tool cycles.

WHO THIS IS FOR

Chief executives, chief operating officers, chiefs of staff, heads of function, and operators accountable for a workflow an AI tool is proposing to change. Whoever will be answerable for this workflow one year from now should be in the room.

WHEN TO USE IT

Use it when a tool is on the table and a purchase decision is active. Use it before the negotiation call, before a pilot, and again at renewal.

WHAT DECISION THIS HELPS MAKE

Whether this tool earns a place in the operating system of the business.

HOW TO RUN A 45 TO 60 MINUTE DECISION REVIEW

Block a meeting with the workflow owner, the executive accountable for the function, and two people who can speak for cost and for data access. Walk the four gates in sequence and mark each one Clear, Concern, or Blocker. Leave the meeting with a written decision, an owner, and the next review point.

FOUR DEFENSIBLE OUTCOMES · EACH IS OPERATIONAL

Decision outcomes

OUTCOME · 01

Buy now.

The workflow is named, owned, and understood. The data is reachable. Review and approval logic is clear. The net gain, after setup, review, training, and coordination, is material. The team can name the work that stops elsewhere. If the vendor were gone in twelve months, the business would feel inconvenience rather than damage.

OUTCOME · 02

Run a bounded pilot.

The workflow is believable. One or two assumptions about operating fit or real gain remain unproven. A narrow, dated test is the cheapest way to retire those unknowns.

OUTCOME · 03

Wait.

The workflow is not ready, the data is not reachable without remediation, or the accountable owner is missing. The right next step is operating work, not procurement.

OUTCOME · 04

Reject.

The purchase is being driven by pressure, narrative, or a polished demo, and the workflow will not survive contact with the tool. Rejection is a legitimate outcome, and often the cheapest one.

EACH GATE TAKES ONE STATUS AT THE END OF THE CONVERSATION, NOT FIELD BY FIELD.

Decision standards

CLEAR

No open questions that would change the decision. What remains is execution detail.

CONCERN

Open questions that can be retired within a pilot, a short diagnostic, or a focused piece of operating work. A Concern shapes the decision; it does not stop it.

BLOCKER

The gate cannot be passed at the current state of the workflow, the environment, or the organization. A Blocker must be resolved upstream of the purchase, not by the purchase.

HOW STATUSES TRANSLATE TO DECISIONS

Gate status combination	Default decision
All four Clear	*Buy now*
Three Clear, one Concern	*Buy now, with conditions*
Two or more Concerns, no Blockers	*Run a bounded pilot*
Any Blocker on Gate 1 or Gate 4	*Wait or Reject*
Blocker on Gate 2 or Gate 3, not resolvable in ninety days	*Reject*
Blocker resolvable by internal operating work within a quarter	*Wait*

> *Treat the table as a starting position. Override it in writing if the decision differs, and record the reason in the final decision record.*

1

Define the workflow.

Confirm the team is buying a workflow change, not a software story.

Answer from today's operating reality, not from the vendor's framing. If a field is unknown, write 'unknown'. Unknowns are the point of this gate. If the workflow does not yet exist in a nameable form, the gate is a Blocker.

Workflow name

Workflow owner (one named person)

Trigger (what starts this workflow)

FREQUENCY AND VOLUME

Runs per day, week, or month (circle the unit)

Typical volume per run (records, cases, documents)

Peak volume and cadence

UPSTREAM DEPENDENCIES

Dependency	Owner	Reliable today?

DOWNSTREAM DEPENDENCIES

Consumer	What they receive	What breaks if wrong

FAILURE COST

What it costs the business when this workflow goes wrong

How damaging one bad output can be ☐ Low ☐ Medium ☐ High ☐ Severe

CURRENT STEP MAP

Judgment: Low for routine, Medium for structured judgment, High for irreducibly human decisions.

#	Step	Role	System	Min	Judgment
1					
2					
3					
4					
5					
6					
7					
8					

INPUTS

Input	Source	Format	Owner

OUTPUTS

Output	Destination	Consumer	Quality bar

WHERE HUMAN JUDGMENT CURRENTLY MATTERS

Name steps, not people. Judgment includes interpretation, discretion, and trade-offs that cannot be reduced to a rule.

1.

2.

3.

STANDARDIZED VERSUS EXCEPTION-HEAVY

Share of runs that follow the standard path _____ percent

Common exception types

1.

2.

3.

How exceptions are handled today

SHARED STANDARD OF CORRECTNESS

Definition of a correct output today

Where that standard is written

Two reviewers rating the same output independently would agree

☐ Mostly yes ☐ Mostly no ☐ Untested

> *Without a shared standard, a tool produces volume against no agreed target, and review multiplies disagreement.*

CURRENT TIME COST PER CYCLE

Minutes per run

Runs per week

Total weekly minutes

CURRENT COORDINATION LOAD

People involved

Handoffs per cycle

Meetings or messages per cycle

TARGET WORKFLOW AFTER THE TOOL

Describe the workflow as it will exist once the tool is in place. Name the steps, the owner, the review points, and the outputs. If this cannot be described in six sentences, the target is not clear enough to buy.

WHAT EXACTLY CHANGES, AND FOR WHOM

Change	Affected role	New behavior

IS THIS ACTUALLY AN AI TOOL PROBLEM?

Some workflows are not waiting for a model. They are waiting for a written standard, cleaner inputs, a clearer handoff, a policy fix, or a single owner.

- ☐ Needs a written standard of correctness before any tool can help.
- ☐ Needs cleaner or more reliable input data before any tool can help.
- ☐ Needs a clear owner or decision right before any tool can help.
- ☐ Needs a policy or governance change before any tool can help.
- ☐ Already well designed; the remaining friction is the kind an AI tool can reduce.

> *If any of the first four boxes is checked, the tool is being asked to carry organizational work it cannot carry. Address the upstream item before buying.*

GATE 1 STATUS

☐ Clear ☐ Concern ☐ Blocker

Notes

2 Test operating fit.

Check whether the tool can function inside the real operating environment.

This gate fails silently more than any other. A tool that cannot reach the data, cannot write back to the system of record, or cannot enforce the right approval will still demo well. Answer for what the tool can do inside this company within the next month.

DATA REACHABILITY

Data	Location	Reachable?	Remediation

DATA SENSITIVITY

Classify the data the tool will touch.

- ☐ Public
- ☐ Internal only
- ☐ Customer personal data
- ☐ Regulated or contractually restricted (health, financial, legal, sector-specific)

Owner of the classification decision

RETENTION, DELETION, TRAINING-DATA USE

Vendor retention of inputs and outputs

Deletion options

Vendor uses company or customer data to train its models

☐ No ☐ Yes, opt-out available
☐ Yes, no opt-out ☐ Unknown

Documented where (contract, DPA, vendor page)

SECURITY, COMPLIANCE, AND RESIDENCY

Vendor certifications (for example, SOC 2, ISO 27001, sector-specific)

Geographic location of the data

Matches contractual and regulatory requirements

☐ Yes ☐ No ☐ Partial

If a customer or regulator asks in writing where their data went, the company answer is:

PERMISSIONS AND ACCESS

Who grants access

Which systems

Which approvals

Expected time to access

SYSTEM OF RECORD

Authoritative system for the output

Tool writes back

☐ Yes ☐ No ☐ Partial

If partial, what does not write back

MANUAL BRIDGE WHEN WRITE-BACK IS INCOMPLETE

If the tool cannot write back in every case, describe the manual bridge.

Bridge step	Owner	Frequency	Time per cycle

> *Without a bridge, the system of record drifts. Concern at minimum, Blocker for regulated workflows.*

DUPLICATE ENTRY AND SWIVEL-CHAIR WORK

Steps that require a person to carry information between the tool and another system by hand

Minutes per cycle

SHADOW SYSTEM RISK

Risk that the tool becomes a second source of truth, parallel to the system of record

☐ Yes ☐ No

If yes, how that will be prevented

HUMAN REVIEW

Steps that require review before release

Who reviews

Average review time per cycle

APPROVAL LOGIC

Who approves the output

Under what conditions

What gets escalated

EXCEPTION PATHS

When the tool produces a wrong or low-confidence output, what happens

Who catches it

How it is logged

AUDITABILITY

A reviewer can reconstruct why the tool produced a given output six months from now

☐ Yes ☐ No ☐ Partial

Where the record is kept

ROLLOUT SCOPE

Initial scope (team, workflow, region)

Expansion criteria

WHO MUST CHANGE BEHAVIOR

Role	Today	After

GATE 2 STATUS

☐ Clear ☐ Concern ☐ Blocker

Notes

3 Test the real gain.

Make sure the gain is real once setup, review, training, and coordination are counted.

The core arithmetic is on a time basis. Money sits in a separate view further down. If the amortized net time gain is close to zero, do not buy.

TIME SAVED

Minutes saved per run

Runs per week

Gross weekly minutes saved

COORDINATION REMOVED (OPTIONAL)

Weekly coordination minutes removed

If coordination does not meaningfully change, enter 0.

COORDINATION ADDED

Weekly coordination minutes added

REVIEW ADDED

Minutes of review per reviewed run

Share of runs that are reviewed percent

Effective review minutes per run (minutes × share)

Weekly review minutes added

WEEKLY NET GAIN, BEFORE AMORTIZATION

Gross weekly minutes saved
Plus weekly coordination removed
Minus weekly review minutes added
Minus weekly coordination added
Equals weekly net gain, before amortization

AMORTIZED WEEKLY BURDEN

Convert first-year setup, training, and ongoing admin into a weekly time cost, spread across fifty working weeks.

Burden	Hours in year 1	Weekly minutes
Internal owner setup		
Training (all users)		
Ongoing admin and configuration		
Total		

WEEKLY NET GAIN, AFTER AMORTIZATION

Weekly net gain, before amortization
Minus total amortized weekly burden
Equals weekly net gain, after amortization

WORK MOVING UPWARD IN THE ORG CHART

Many tools remove work from a lower-cost role while adding review or exception work to a scarcer, higher-cost role. Map it before buying.

Role losing work	What comes off	Role gaining work	What gets added

> *If the role gaining work is the bottleneck role, the business gain is smaller than the time arithmetic suggests.*

WHERE QUALITY IMPROVES, STAYS FLAT, OR GETS WORSE

Fill all three columns for every dimension. An empty 'gets worse' column is a sign the review has been too generous.

Dimension	Improves	Flat	Gets worse
Accuracy			
Speed			
Consistency			
Auditability			
Customer experience			
Reviewer workload			
Exception handling			

WHICH ROLE CAPTURES THE GAIN

Named role

What they will do with the reclaimed capacity

If the gain has no named owner and no named use, it will not be captured. Write 'unassigned' if that is the truth.

WHAT THE TEAM WILL STOP DOING

Name the specific activities that come off the plate because of the tool. If this table cannot be filled, the tool is being added on top of existing work.

Activity stopping	Owner	When

An empty table is a Concern at minimum.

DIRECT SPEND · money, separate view

Annual tool cost (subscription, per seat, usage) _____

One-time implementation cost _____

Year-one total spend _____

OPTIONAL PAYBACK SANITY CHECK

Annual net time saved (weekly net gain after amortization × 50) _____ hours

Year-one total spend _____

Priced at a reasonable loaded hourly rate _____

Payback period (months) _____

> *Use payback as a sanity line. Under twelve months is a reasonable purchase when Gate 4 is Clear, and a trap when Gate 4 is a Blocker.*

GATE 3 STATUS

☐ Clear ☐ Concern ☐ Blocker

Notes _____

4

Test survivability.

Make sure the organization is not fragile if the vendor stalls, raises prices, or disappears.

As a planning exercise, assume the vendor could be acquired, raise prices sharply, or shut the product down within eighteen months. The goal is to know the organization's exposure.

EXPORTABILITY

The tool's outputs, configurations, prompts, and data can be exported in a usable format

☐ Yes ☐ No ☐ Partial

Export format

Owner of the export

Cadence

FALLBACK WORKFLOW

If the tool were gone tomorrow, what the team would do

Time per cycle on the fallback

SWITCHING COST

Time to move to an alternative

Cost to move

Data or configuration that would not transfer

RETAINED VERSUS TRAPPED KNOWLEDGE

If the trapped list is longer than the retained list, the team is building fragility.

Retained (in documents, playbooks, systems of record, shared team practice)

Trapped (only inside the tool, its prompts, its configuration, or one person's head)

SINGLE-PERSON DEPENDENCY

Who, if they left, would take critical tool knowledge with them

What would be lost

Plan to document it

SINGLE-VENDOR DEPENDENCY

A second vendor can run this workflow within a quarter

☐ Yes ☐ No

Named alternative

What it would take to qualify them

WHAT BREAKS IN THE FIRST WEEK IF THE TOOL IS GONE

1.
2.
3.
4.

THE TWELVE-MONTH DISAPPEARANCE TEST

Assume twelve months from now, the tool has been gone for a week.

Workflow still runs ☐ Yes ☐ No ☐ Degraded

Customers are lost ☐ Yes ☐ No

A role becomes impossible ☐ Yes ☐ No

Written knowledge for the team to fall back on

☐ Yes ☐ No

> *Three or more answers pointing to fragility make Gate 4 a Blocker until a fallback is designed.*

GATE 4 STATUS

☐ Clear ☐ Concern ☐ Blocker

Notes

THE RECORD · 05

Bounded pilot design

Use this only if the decision is to run a pilot. A pilot is a dated test designed to retire specific unknowns. It is not a soft purchase, and not an open-ended trial.

SPECIFIC UNKNOWNS THE PILOT WILL RETIRE

1.
2.
3.

PILOT SCOPE

In scope

Explicitly out of scope

WORKFLOW SEGMENT BEING TESTED

USERS

Number

Named users

Why these users

OWNER

Pilot owner

Accountable executive

DATES

Start

Midpoint review

End

Decision date

BASELINE MEASUREMENT BEFORE THE PILOT BEGINS

Without a baseline, the pilot cannot produce a comparison.

Metric	Baseline	How measured	Owner	Date

SUCCESS CRITERIA

Specific, observable, numeric where possible.

1.
2.
3.

FAILURE CRITERIA

If any is hit before the end date, the pilot ends early.

1.
2.
3.

INVALIDATION CRITERIA · separate from failure

Conditions under which the pilot produces no usable signal, regardless of results.

1.
2.
3.

REVIEW CADENCE

How often the pilot group meets

What gets reported at each review

MAXIMUM TOLERATED WORKAROUND BURDEN

Maximum hours per week of workarounds

Maximum open workaround items

Escalation path if the threshold is crossed

ROLLBACK AND EXIT PLAN

How the workflow returns to its state before the pilot if the pilot ends

Owner

Time required

SCOPE AND CONVERSION DISCIPLINE

Pilot scope, users, and workflow segment are fixed for the duration. Any expansion requires a written scope-change approval by the accountable executive, and resets the midpoint review.

> *Pilot success does not automatically convert to a purchase. A Buy decision requires a fresh review against the four gates, on or after the decision date.*

WHAT EVIDENCE WILL COUNT

Time-tracked samples, reviewer error logs, user diaries, system logs, customer outcomes. Name the sources before the pilot starts.

1.
2.
3.

WHAT WILL BE IGNORED AS NOISE

Anecdotal enthusiasm, vendor-produced metrics, case studies from other companies, a single champion's testimonial.

1.
2.
3.

THE RECORD · 06

Final decision record

The one page that leaves the meeting. Everything else is supporting material.

Tool name

Vendor

Workflow

GATE SUMMARY

Gate	Status	Reason
1. Workflow	☐ Clear ☐ Concern ☐ Blocker	
2. Operating fit	☐ Clear ☐ Concern ☐ Blocker	
3. Real gain	☐ Clear ☐ Concern ☐ Blocker	
4. Survivability	☐ Clear ☐ Concern ☐ Blocker	

DECISION

☐ Buy now ☐ Pilot ☐ Wait ☐ Reject

DECISION OWNER

Named person accountable for the outcome

RATIONALE · four to six sentences, written in the meeting

CONDITIONS TO PROCEED

Condition	Owner	Due

REVIEW TRIGGERS

Developments inside the next quarter that would reopen the decision.

1.

2.

3.

NEXT REVIEW POINT · mandatory if Buy now

Date

Agenda

SIGNATURES

Accountable executive

Date

Finance

Date

Workflow owner

Date

Data and security

Date

THE RECORD · 07

Red flags and stop signs

If any of the following is true, stop the purchase until it is resolved.

01 **Unclear workflow owner.**
No single named person is accountable. A team or committee cannot be the owner.

02 **No reachable data.**
The tool needs data it cannot access without a separate engineering project. Remediate the data first.

03 **No baseline.**
Current performance has not been measured. Any improvement claim after purchase cannot be verified.

04 **Demo quality mistaken for production fit.**
The decision is being driven by a polished demo on clean data. The gain rarely transfers to production conditions.

05 **Hidden review burden.**
Time savings look good only because review has not been costed. AI outputs that matter will be reviewed by a human.

06 **No fallback if the tool is gone.**
The team cannot describe what it would do in the first week without it.

07 **Work added, not replaced.**
The team cannot name what comes off the plate. Adding without subtracting slows the operating system down.

08 **Pressure from outside the operating team.**
The purchase is being pushed by a board member, an investor, a peer company, or the vendor, rather than by the workflow owner.

09 **Single-person dependency.**
One person holds the configuration, prompts, and reviews. That concentrates risk.

10 **Workflow named but not designed.**
The team can name the workflow but cannot describe the steps before and after the tool. Name is not design.

11 **Workflow volume too low or irregular.**
The workflow runs a few times a month, or in unpredictable bursts. The setup cost will not amortize.

12 **Upstream mess mistaken for an AI opportunity.**
The real problem is bad data, unclear handoffs, or missing standards upstream. A tool placed on top will inherit them.

13 **No shared standard of correctness.**
Reviewers cannot agree on what a correct output looks like. A tool producing outputs at volume against no standard multiplies disagreement.

14 **Required behavior change exceeds likely adoption.**
The tool only delivers if a meaningful share of the team changes how they work, and the organization lacks the capacity to drive that change.

15 **Workflow too exception-heavy for the proposed gain.**
The standard path is a small share of total volume. The gain from automating it does not justify the cost.

CLOSING NOTE · 08

> *A defensible reject is as valuable as a defensible buy. Both are written decisions that hold up when read six months later by someone who was not in the room. What matters at twelve months is whether the workflow is faster, the team knows why, and the vendor can be replaced without disruption.*