



AI business case and ROI

Building an honest case for AI investment, with baselines, costs, benefits and the evidence to stop or scale.

WORKS WITH THE ASSISTANTS TEAMS ALREADY USE



EDITION 37 / EMBER & MINT

EXECUTIVE WORKING EDITION - 08 OCT 2026

AT A GLANCE

Building an honest case for AI investment, with baselines, costs, benefits and the evidence to stop or scale.

Most AI business cases fail in one of two ways. They count only the license fee and promise large savings that never show up, or they are so cautious that nothing gets funded. This playbook shows how to build a case that survives a finance review: a measured baseline, a full cost model that includes review time, change effort, data work and ongoing evaluation, benefits sorted by how hard they are to prove, and stage gates that release money only when evidence arrives. It ends with a stop-or-scale review that treats stopping as a legitimate and useful result.

WHO IT IS FOR

Finance partners, business unit leaders, transformation and operations leaders, and sponsors who must approve, fund or defend an AI investment, together with the analysts who build the case.

TIME TO USE

75 minutes to read; two working sessions to build a first case

What you will have at the end

- A baseline for the process you want to change, measured before any tool is bought
- A full cost model with the hidden costs made visible
- Benefits classified by evidence strength, with illustrative maths you can adapt
- Stage gates with explicit criteria for release of funds
- A stop-or-scale review format that your finance team will trust

Contents

01	Why AI business cases go wrong	03
02	Frame the investment around a process and a decision	04
03	Measure the baseline before you buy anything	05
04	Build the full cost model, including the hidden costs	07
05	Classify benefits by how well you can prove them	09
06	Honest ROI maths with a worked illustration	10
07	Stage gates: release money as evidence arrives	12
08	Track benefits after launch and run stop-or-scale reviews	13
09	Common pitfalls and how to present the case	15
10	A first quarter that produces evidence	16
11	Terms used in this publication	17

EXECUTIVE BRIEF

Why AI business cases go wrong

01



The usual failure is not optimism alone. It is a case built on a tool price and a hoped-for percentage, with no baseline and no owner for the benefit.

A typical weak case reads like this: the tool costs a modest amount per user, staff spend many hours on a task, the tool will save a fraction of that time, therefore the return is large. Every link in that chain is untested. The hours were estimated, not measured. The fraction comes from a vendor slide. The saved time is assumed to turn into money, which it rarely does unless someone decides what happens to the freed capacity. Finance teams have seen this pattern for decades with other technology, and they discount it accordingly.

An honest case does something different. It starts from a process, not a product. It measures how that process performs today in time, cost, quality and volume. It lists every cost, including the ones nobody wants to mention. It separates benefits that will appear in the accounts from benefits that are real but soft. And it commits to a test with a pass mark before the larger spend is released. The result is a smaller number that people believe, which is worth more than a large number they do not.

- Cases fail when the benefit has no owner who will carry it in a budget
- Cases fail when time saved is counted as money saved without a decision about the freed time
- Cases fail when costs stop at the license and ignore the people effort around it
- Cases fail when there is no baseline, so improvement cannot be shown even if it happens
- Cases fail when the pilot has no pass mark, so every result is read as success

PLAIN TEST

Could a skeptical finance partner, given only your baseline data and cost model, rebuild your return figure and arrive at a similar number? If not, the case is not ready to present.



WORKING QUESTION

If your sponsor had to explain this investment to the finance committee in three sentences, which sentence would be hardest to defend with data?

FRAMING

Frame the investment around a process and a decision

02

Fund a change to a named process with a named owner, not a general intention to use AI.

The unit of investment should be a process outcome such as faster quote turnaround, fewer rework loops in claims handling or shorter time to first useful draft in proposal writing. A process has a volume, a cost per unit, a quality standard and a person accountable for it. These are the handles a business case needs. A statement such as we will give everyone an assistant is a license purchase, and it should be argued as one: low cost, low expected return, limited downside, reviewed after a fixed period.

Write the case as a decision. State what is being decided, by whom, by when and what happens if the answer is no. Name the alternatives, including doing nothing and improving the process without AI. Many processes have a simpler fix such as removing an approval step or cleaning an input form, and an honest case compares against that option. If AI is not the best way to capture the value, say so; the credibility gained is worth the lost project.

INVESTMENT TYPE	TYPICAL SHAPE	HOW TO ARGUE IT
Broad productivity license	Many users, small individual gains hard to measure	Low cost, time-boxed, judged on adoption and sampled time studies
Process redesign with AI	One workflow, measurable volume and cost per unit	Full case with baseline, costs, benefits and gates
New capability	Something not done today at any price	Option value; stage gates; explicit upper limit on spend

INVESTMENT TYPE	TYPICAL SHAPE	HOW TO ARGUE IT
Risk or quality reduction	Fewer errors, fewer incidents	Expected loss avoided; use cautious, labelled estimates
Platform or data foundation	Enables several later use cases	Argue on the use cases it unlocks, each with its own case

Case frame, one page

- Process named, with an accountable owner
- Decision, decision maker and date written down
- Current performance measured, not estimated
- Alternatives listed, including no AI and do nothing
- Success measures and pass mark agreed before the pilot
- Budget limit for each stage stated

WORKING QUESTION

Which single process, with which accountable owner, would carry the benefit if this investment worked?

BASELINE

Measure the baseline before you buy anything

03

Without a before picture there is no honest after picture. A baseline takes days, not months, and it is the cheapest evidence you will ever buy.

A baseline records how the process performs now, using the same measures you will use later. For a document process that usually means volume per week, elapsed time from request to delivery, touch time per item, rework rate, error or escalation rate and the cost of the people involved. Touch time and elapsed time are different. An item may wait three days in a queue

and need forty minutes of actual work. AI usually changes touch time first, while elapsed time depends on queues and approvals that AI does not touch.

01 Map the process

Walk the steps with the people who do the work. Note inputs, hand-offs, approvals and where rework loops occur.

02 Sample real items

Pull a representative sample, for example thirty to fifty recent items, including difficult ones. Do not use only the easy cases.

03 Time the work

Have a few staff record touch time on live items for one to two weeks. Self-reported estimates from memory tend to be rough, so prefer recorded times.

04 Score quality

Define what good looks like and have a second reviewer score the sample, so quality can be compared later.

05 Agree the numbers

Share the baseline with the process owner and finance, and record it as the reference point with a date.

BASILINE MEASURE	WHY IT MATTERS	COMMON TRAP
Volume per period	Benefits scale with it	Peaks and seasonality hidden in an average
Touch time per item	Where AI usually helps first	Estimates from memory instead of recorded times
Elapsed time	What customers and colleagues feel	Assuming AI shortens queues it does not touch
Rework and error rate	Quality side of the case	Not measured, so quality drift goes unseen
Fully loaded labor cost	Converts time to money	Using salary alone and ignoring overhead, or the reverse
Escalation rate	Shows where judgment is needed	Ignoring it, then wondering why savings stall

COMMON MISTAKE

Letting the vendor supply the baseline. A vendor benchmark describes other organizations and their processes. Your case needs your numbers, collected by your people, agreed by your finance partner.

K

WORKING QUESTION

For the process you have in mind, which of the baseline measures could you fill in today from recorded data rather than opinion?

COST MODEL

Build the full cost model, including the hidden costs

04

License fees are usually the smallest and easiest part. The costs that sink returns are people time, data work and the permanent cost of keeping the system correct.

A cost model has three layers. Build cost covers what you spend to get live: tool setup, integration, data preparation, security review, legal review and design. Change cost covers training, new routines, communication and the productivity dip while people learn. Run cost covers what recurs: licenses or usage fees, hosting, support, monitoring, evaluation, content upkeep and the human review that remains in the process. Treat run cost as a multi-year commitment, because it continues as long as the system is in use.

COST CATEGORY	WHAT IT INCLUDES	OFTEN MISSED BECAUSE
Licenses and usage	Seats, usage-based fees, platform charges	Usage can grow faster than seats; model it at several volumes
Review time	Human checking of outputs before use	Treated as free, yet it can consume much of the saved time

Ø

COST CATEGORY	WHAT IT INCLUDES	OFTEN MISSED BECAUSE
Data work	Cleaning, labelling, structuring, access setup	Seen as an IT task and left unbudgeted
Integration	Connecting to systems of record and identity	Estimated from a demo rather than your systems
Change effort	Training, champions, manager time, communication	Not seen as a cost because it uses existing staff
Ongoing evaluation	Test sets, sampling, monitoring, regression checks	Assumed to end at launch
Governance and risk	Reviews, audits, legal and security input	Counted only once, not annually
Exit and switching	Export, migration, retraining	Ignored until the vendor changes terms

Review time deserves its own line

If a person must read and correct every AI draft, the net saving is the time to produce the draft minus the time to review it, plus any extra time spent fixing errors that a first-draft human would not have made. Early in a rollout the review burden is usually highest, because people do not yet know where the system is reliable. Model review time as a share of the original task, state your assumption, then measure it in the pilot and replace the assumption with data.

ILLUSTRATIVE EXAMPLE ONLY

Suppose a task takes 40 minutes by hand. A draft takes 5 minutes to request and 15 minutes to review and correct, so the net saving is 20 minutes, or half. A case that assumed 35 minutes saved would overstate the benefit by 75 percent. All figures are hypothetical.

WORKING QUESTION

Which cost line in your draft model is currently zero or blank, and who could tell you what it really is?

BENEFITS

Classify benefits by how well you can prove them

05



Not every benefit is equal. Sort them by evidence strength and count only what you can defend.

Benefits fall into recognizable types. Cash benefits appear in the accounts: lower external spend, avoided hires against a stated plan, reduced overtime, fewer refunds. Capacity benefits free time that is redeployed to work with its own value, such as more customer conversations. Quality and risk benefits reduce errors, rework and exposure. Strategic benefits improve speed to market or customer experience. The first type is easiest to prove and the last is hardest, but the last can be the most valuable. A good case shows all four and is clear about which are counted in the return figure.

BENEFIT TYPE	EXAMPLE	EVIDENCE NEEDED	COUNT IN ROI?
Cash	Lower agency or contractor spend on drafting	Budget line before and after, owner sign-off	Yes
Capacity, redeployed	Analysts handle more accounts with the same team	Volume per person up, with output tracked	Yes, at a stated discount
Capacity, unassigned	Minutes saved spread across many people	Time studies only	Show separately, not in headline
Quality and risk	Fewer rework loops, fewer escalations	Sampled quality scores, error counts	Cautiously, labelled estimates
Strategic	Faster response to customers	Cycle time and win or retention data	Narrative, not in headline



The key discipline is the redeployment decision. If each of forty people saves ten minutes a day, no cash appears unless the organization decides to do something with that time: take on more volume, reduce temporary staffing, delay a hire or shift people to higher-value work. Write that decision into the case, name who makes it and set a date. Where the answer is that the time will be absorbed into the day with no change, count the benefit as capacity that improves experience, not as money.



- Name a benefit owner for each line, someone whose budget or target moves if the benefit is real
- Apply a realization factor to capacity gains, because not all saved time is usable
- Avoid double counting the same hour as both cash and capacity
- Keep soft benefits visible but outside the headline return
- Record the date and method by which each benefit will be verified

WORKING QUESTION

For each benefit in your case, who would you ask to confirm it in six months, and what number would they point to?

S.
ROI MATHS

Honest ROI maths with a worked illustration

06



Use simple arithmetic, state every assumption, and show a range. A conservative case that holds up beats a bold one that collapses under questions.

Three measures serve most decisions. Net benefit is total benefits minus total costs over the period. Return on investment is net benefit divided by total cost. Payback period is the time until cumulative benefits cover cumulative costs. Use a period that matches reality, usually three years for tools whose terms and capabilities change, and discount later years if your finance team does. Always show a low, expected and high scenario, and say what has to be true for each.

K.

Worked illustration (all numbers hypothetical)

A claims team handles 2,000 routine items a month. Baseline touch time is 30 minutes. A pilot suggests AI drafting cuts it to 18 minutes including review, a saving of 12 minutes, or 40 percent. That is 400 hours a month. The fully loaded cost of an hour is set at 45 in this example. The gross value is 18,000 a month. Because only some of the freed time can be redeployed, apply a 60 percent realization factor, giving 10,800 a month of counted benefit.

LINE (ILLUSTRATIVE)	YEAR 1	YEAR 2	YEAR 3
Counted benefit (10,800 a month)	129,600	129,600	129,600
Licenses and usage	30,000	32,000	34,000
Build, data work and integration	45,000	5,000	5,000
Change effort and training	20,000	8,000	6,000
Evaluation, monitoring and content upkeep	12,000	12,000	12,000
Total cost	107,000	57,000	57,000
Net benefit	22,600	72,600	72,600

Over three years in this example, benefits are 388,800 and costs are 221,000, for a net benefit of 167,800 and a return of about 76 percent on total cost. Payback arrives in the second year. Now stress it. If the saving is 8 minutes instead of 12, benefits fall by a third and year one turns negative. If volume drops by a quarter, the same happens. Showing this sensitivity is not weakness. It tells the sponsor which assumption to protect and which to test first.

REMINDER

These numbers are invented to show the method. They are not benchmarks, and no real organization should expect them. Replace every figure with your baseline and pilot data.

WORKING QUESTION

Which single assumption in your own model, if it were wrong by a third, would flip the case from positive to negative?

STAGE GATES

Stage gates: release money as evidence arrives

07

Fund the question, not the whole answer. Each gate has criteria set in advance and a decision maker who can say no.

Stage gates convert uncertainty into a series of smaller bets. The first stage spends little to learn whether the idea works on your data. The next spends more to learn whether people use it and whether the benefit shows up in real work. Only then does the large spend for scale begin. Each gate lists the evidence required, the budget released and the result that would stop the project. Gates also protect the team, because a pre-agreed stop criterion turns a failed pilot into a planned outcome rather than a failure.

01 Gate 0: Frame

Process, owner, baseline and success measures are agreed. Small budget for discovery and test design. Decide whether the idea is worth testing at all.

02 Gate 1: Prove it works

A scripted test on your own data meets the quality pass mark, with risks and data access understood. Release pilot funds.

03 Gate 2: Prove it is used and pays

A live pilot with real users shows adoption, measured time and quality against baseline, and a realistic review burden. Release rollout funds for one area.

04 Gate 3: Scale

Benefits verified by the benefit owner, run costs known from actuals, support model in place. Release funds for wider deployment in tranches.

05 Gate 4: Sustain

Annual review of value, cost, risk and alternatives. Renew, change or retire.

GATE	EVIDENCE REQUIRED	TYPICAL STOP SIGNAL
Gate 1	Quality pass mark met on a test set from real work	Quality below the mark after two tuning rounds

GATE	EVIDENCE REQUIRED	TYPICAL STOP SIGNAL
Gate 2	Active use by most of the pilot group; measured time saved	Use drops after the novelty fades
Gate 2	Review time within modeled range	Review time consumes most of the saving
Gate 3	Benefit owner signs off on realized benefit	Benefit visible only in surveys, not in work data
Gate 4	Value exceeds run cost with margin	Cost rising faster than value

COMMON MISTAKE

Setting the pass mark after seeing the results. Write the thresholds into the approval paper before the pilot starts, and keep the paper.

WORKING QUESTION

What result in the pilot would make you stop, and have you written it down where the sponsor can see it?

TRACKING

Track benefits after launch and run stop-or-scale reviews

08

Return is realized in operations, not in approval meetings. Build a benefits register and review it on a fixed rhythm.

After launch, the business case becomes a benefits register: each benefit line, its owner, the measure, the baseline, the target, the current value and the date of next review. Update it monthly during the first six months and quarterly after that. Pair it with a cost register fed by actual invoices and time records, so the return is calculated from what happened and not from what was planned. Where the figures diverge from the case, the review asks why and decides

what to do.

REVIEW QUESTION	EVIDENCE TO BRING	POSSIBLE DECISION
Is it used?	Active users, frequency, task coverage	Retrain, simplify or retire
Is it good?	Sampled quality scores and error trends	Tune, restrict scope or add review
Is it faster or cheaper?	Touch time, cost per unit against baseline	Scale, or revise the benefit target
Is the benefit banked?	Budget movement, redeployment evidence	Confirm, adjust forecast or reclassify
Are costs as planned?	Invoices, usage, support effort	Renegotiate, cap usage or switch
Is there a better option now?	Market changes, simpler alternatives	Replace or continue

- Scale when quality meets the mark, use is stable and the benefit owner confirms the number
- Change when value exists but cost, scope or process design is off
- Pause when evidence is mixed and a specific test could resolve it within weeks
- Stop when value is not appearing, the review burden defeats the saving, or risk exceeds benefit
- Record what was learned in either case, so the next case is stronger

Stopping deserves respect. A project that stops at Gate 2 with clear evidence has cost a fraction of one that drifted for a year. Publish the learning, thank the team and move the capacity to the next candidate. Organizations that treat every stop as a failure get optimistic reporting; organizations that treat it as a decision get honest numbers, and honest numbers are what make the next investment easier to approve.

WORKING QUESTION

Who in your organization has the authority and the incentive to say this should stop, and have you told them what evidence would make them say it?

PITFALLS AND PRESENTATION

Common pitfalls and how to present the case

09

A good case is short, specific and open about what it does not know. Present it as a decision with evidence, not as advocacy.

- Quoting vendor case studies as if they described your process
- Counting minutes saved across hundreds of people as headcount reduction
- Leaving out review time, training time and the cost of keeping content current
- Presenting a single figure with no range and no stated assumptions
- Combining several unrelated use cases into one blended return that no one owns
- Skipping the do-nothing and non-AI alternatives
- Treating risk and compliance cost as a one-time item

ILLUSTRATIVE SCENARIO (HYPOTHETICAL)

A team proposes licenses for 300 staff on a promised saving of one hour per week each. Finance asks for a baseline. The team times a sample and finds the saving applies to about a quarter of the staff and is closer to twenty minutes. The case is rewritten for 80 users in two functions, with a gate at 60 days. It is approved because the numbers are now believable.

Before the case goes to the committee

- Baseline measured, dated and agreed by the process owner
- Cost model covers build, change and run, with review time and evaluation lines
- Each benefit has an owner, a measure and a verification date
- Low, expected and high scenarios are shown with the key assumptions
- Stage gates, pass marks and stop criteria are written down
- Alternatives and risks are stated plainly

- Legal, security and data protection input has been requested, with current requirements checked



Keep the paper to a few pages: the decision, the process and baseline, the cost model, the benefit table, the scenarios, the gates and the risks. Put the detail in an appendix. Lead with what you want approved and for how much at this stage, not the full multi-year number. Invite challenge, and bring the person who owns the benefit to the meeting, because their commitment is the strongest evidence in the room.

WORKING QUESTION

If the committee approved only the first stage, would the evidence from that stage be enough for them to approve or reject the next?



30 / 60 / 90

A first quarter that produces evidence

10



Build the case from measured facts, fund it in stages and let the evidence decide what happens next.

First 30 days: ground it

- Choose one process and name its owner and the benefit owner
- Measure the baseline for volume, touch time, quality and cost
- Draft the cost model with review time, data work, change and evaluation lines
- Write the pass mark and stop criteria for Gate 1

Days 31 to 60: test it

- Run the scripted test on real work and score it against the pass mark
- Replace modeled assumptions with measured ones, especially review time
- Build low, expected and high scenarios and test the sensitive assumptions
- Take the stage-gated case to the approving committee



Days 61 to 90: decide

- Run the live pilot and track use, quality, time and cost against baseline
- Open the benefits and cost registers with named owners
- Hold the Gate 2 review and record the decision and the reasoning
- Plan the next tranche or the orderly stop, and share what was learned



GLOSSARY

Terms used in this publication

11



TERM	MEANING
Baseline	A dated measurement of how a process performs before any change, used as the reference for later comparison.
Touch time	The time people actively spend working on an item, as opposed to time spent waiting in a queue.
Fully loaded cost	The cost of an hour of work including salary, benefits and overhead, as defined by your finance team.
Realization factor	The share of a gross time saving that is actually converted into cash or usable capacity.
Net benefit	Total counted benefits minus total costs over the chosen period.
Return on investment	Net benefit divided by total cost, usually shown with the period and the assumptions behind it.
Payback period	The time until cumulative benefits equal cumulative costs.
Sensitivity analysis	Testing how the result changes when a key assumption is raised or lowered.
Stage gate	A decision point with pre-agreed evidence and budget, at which a project proceeds, changes or stops.
Benefits register	A tracked list of benefit lines, each with an owner, measure, baseline, target and review date.
Review burden	The human time needed to check and correct AI output before it is used.



TERM	MEANING
Pass mark	A quality threshold agreed before a test, which the result must meet for the project to proceed.



A note on use

This publication is a public orientation resource, not legal, technical, financial or professional advice. Adapt it to your people, information, workflows and obligations, and check current tool terms, regulations and risks before acting.

TAKE THE NEXT STEP

Take the free Halden assessment at halden.company/assessments, or talk to us at halden.company/contact about turning this into a plan for your organization.

