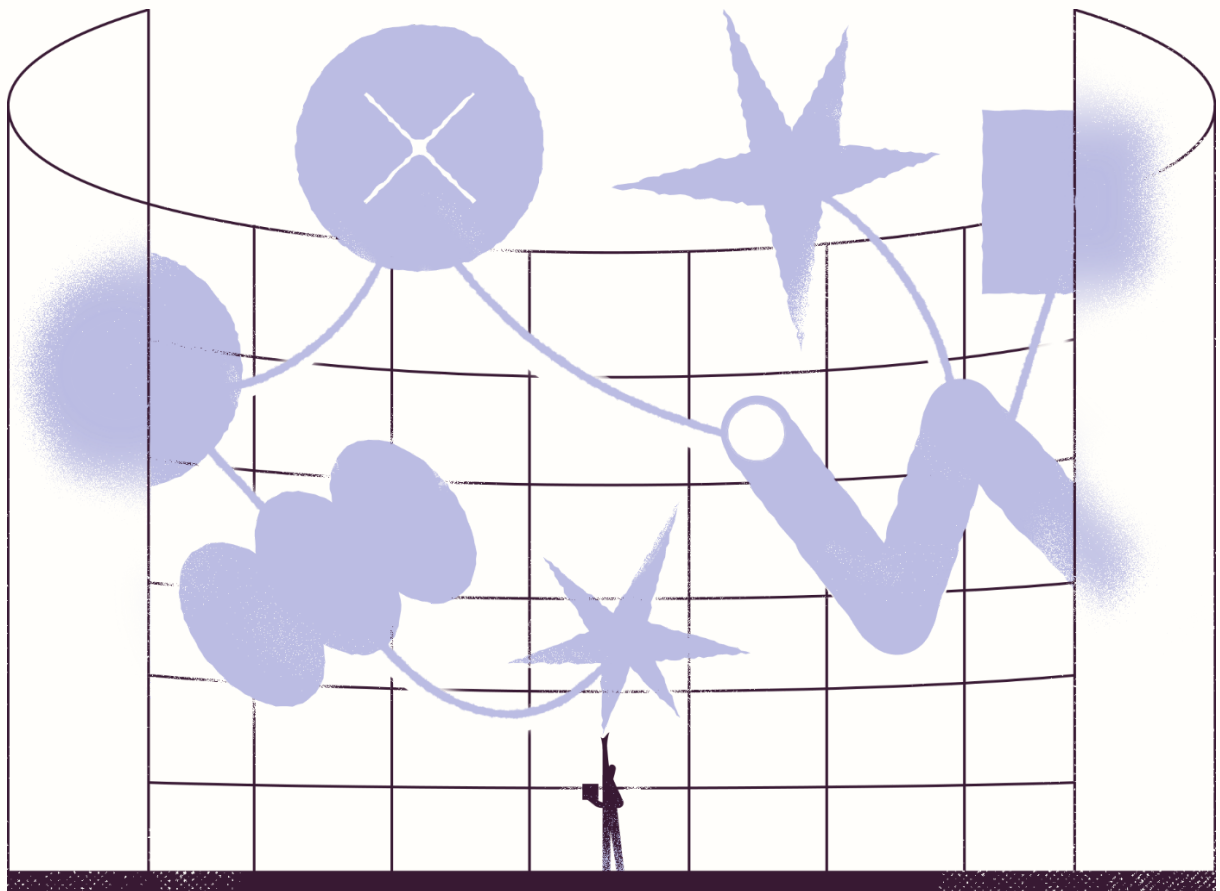


Estimating the Value of AI Automation



Executive summary

Most companies can predict their next month's AI bill, within a certain range. However, the same companies find it challenging to estimate the value of the AI automation work.

The estimation challenge has two parts. Estimating what a use case is worth requires breaking apart value into parts you can source and defend. Estimating what it takes to deliver requires understanding integration complexity, data readiness, and how to deploy AI inside a workflow.

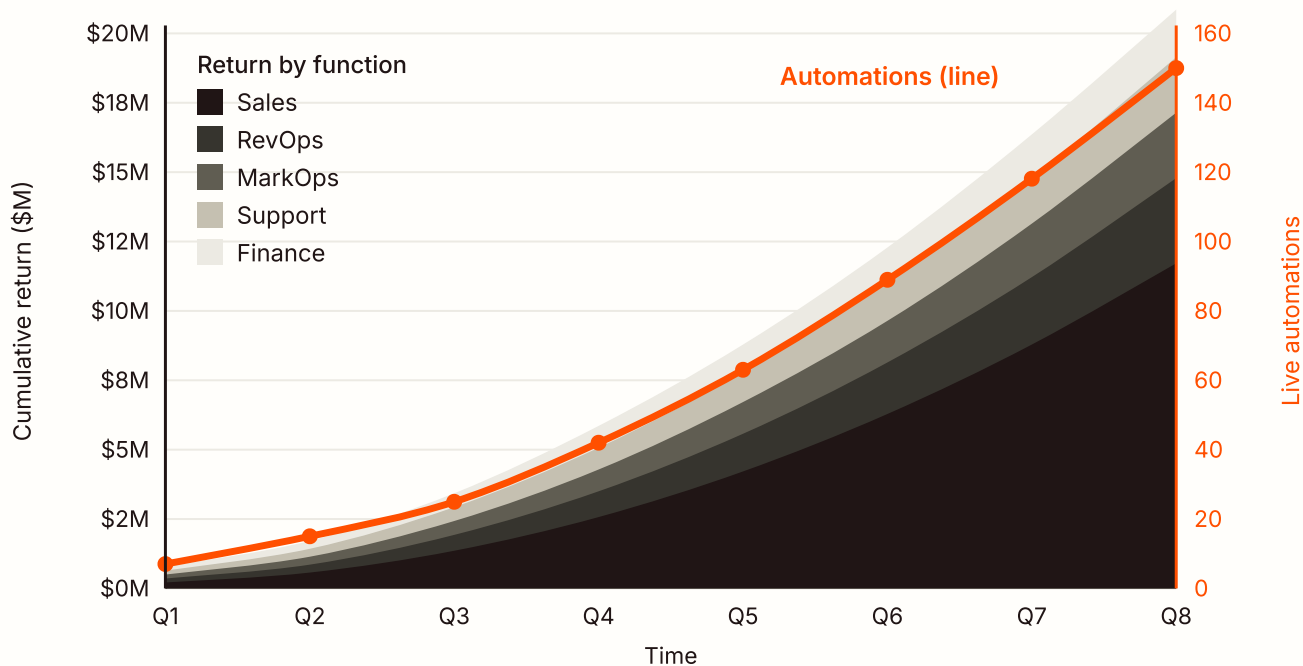
A business case needs both: value alone doesn't tell you whether a project is deliverable, and effort alone doesn't tell you whether it's worth doing. Most broken AI business cases trace back to skipping one half of that pair.

One line of math connects the two:

$$\text{Realizable value} = \text{gross value} \times \text{probability of capture} - \text{cost to build}$$

This paper documents the Unified Value System Zapier uses with enterprise customers to build that case. The effort part gets a multi-factor feasibility rating that feeds a prioritization grid. The value part gets the deeper treatment: five dimensions and seventeen value drivers. The full reference table is in Appendix A. To make the math concrete, this paper walks one sales automation, a lead entered on a web form through to a closed deal, end to end.

Figure 1 shows this in more detail: as more use cases reach production, the return increases.



Illustrative. Return and automation count climb together.

Figure 1. Cumulative return by function as live automations grow. Illustrative.

Value is only half the picture

Every potential use case has two numbers: estimated value and what it will take to build.

The grid below puts both numbers into focus. Workflows in the top-right — high value, easier to deliver — get funded first. The top-left is the second wave: high value on a harder build, worth scoping and scheduling behind the quick wins. The rest of this paper shows how to place a real use case on these two axes with figures that both vendors and buyer champions can explain to a CFO.

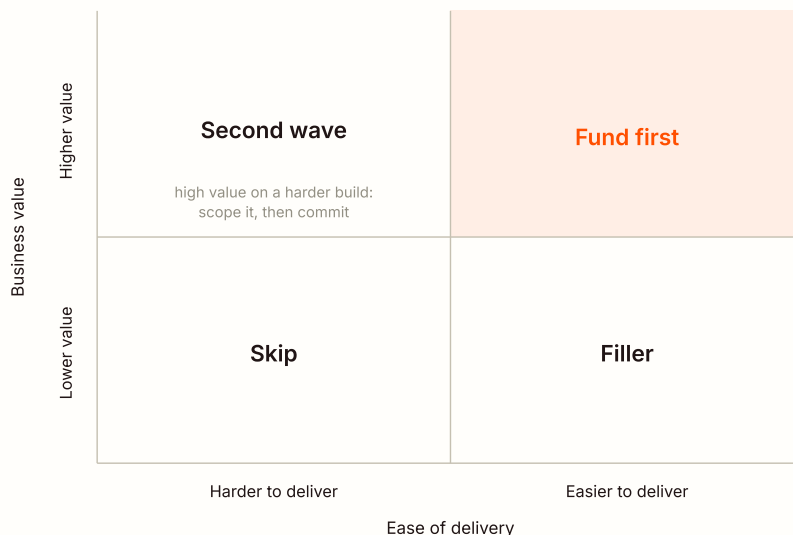


Figure 2. Sort every candidate by value and effort before you fund anything.

The five dimensions of value

Value maps to five dimensions, designed to be mutually exclusive with no overlap and nothing obvious left out. Most use cases register on several dimensions at once, which is why the totals later in this paper add across dimensions. The seventeen underlying value drivers with formula breakdowns are listed in Appendix A.

Revenue Impact — more deals closed with the same sales force, less revenue leaked, faster time to first dollar. This is a top priority for a CRO: capturing more deals without adding headcount. Hardest to attribute, but the most compelling dimension to an economic buyer.

Speed / Cycle Time — the elapsed calendar time a process takes, measured in time, not dollars. The value of the effort that time frees up is counted once, under Productivity.

Productivity — the human effort a process requires. The most commonly quantified dimension, and the easiest for companies and vendors to grasp.

Cost Avoidance — money not spent: a hire not made, a tool not bought, rework not needed.

Risk & Quality — the value of a bad outcome that didn't happen: a violation avoided, data kept clean, an incident prevented. The evidence here is often the hardest to source, so these lines are the ones most likely to be flagged directional or left out, but they are frequently critical to specific stakeholders.

Why the way you build matters

How a workflow is built determines both terms in the equation, so architecture is a value decision, not just a technical one. A deterministic workflow is fully predictable, which increases certainty in the value estimate, but it isn't as extensible as an intelligent workflow, so its potential impact is more limited. A fully agentic workflow is highly extensible but unpredictable, which introduces new challenges in estimating value.

A hybrid workflow uses AI only where it adds something: AI handles the step that needs judgment. Everything else, the lookups, the writes, and the notifications, runs identically every time. That also costs less to run, since AI is only invoked at the judgment step rather than on every action.

Since only the judgment step varies, the rest of the workflow behaves predictably in production, and each fixed step is something you can count.

That combination is what makes hybrid workflows the easiest architecture to build a business case around. It's wide enough in scope to make the use case worth addressing, predictable enough that the value estimate is credible.

A worked example: lead to closed deal

Picture a 5,000-person software company. Without automation, a RevOps person reads each inbound lead, looks up the account, enriches it by hand, decides who owns it, keys it into the CRM, and sends a first reply, and some leads go cold before anyone reaches them.

The automated version reads the message with AI, works out intent and fit, enriches the record from connected sources, routes it to the right rep, writes the CRM record, sends a first reply, and drafts a quote once the lead qualifies. Every number below is illustrative, sized to a ~5,000-person company, and this one use case touches all five value dimensions:

Revenue Impact — faster, cleaner routing converts more of the same leads (Pipeline Velocity).

Speed — first-response time falls from about six hours to a few minutes. Speed measures elapsed time and carries no dollar figure, so the same saved minutes aren't also counted under Productivity, since each saved hour belongs to exactly one driver.

Productivity — Task Elimination is the largest line here, since manual enrichment and data entry stop entirely, which frees sellers to spend their time selling rather than on admin.

Cost Avoidance — cleaner records mean fewer bad ones to chase and re-key later (Error & Rework Elimination).

Risk & Quality — Data Integrity covers what goes wrong when the CRM fills with bad data. Of the five drivers here, its evidence is the weakest, a useful example of how to handle a number you can't fully stand behind (see Appendix B on evidence tiers).

The two CRO priorities we leave unmodeled are better manager visibility, which sharpens coaching, and higher quota attainment, which helps retain top reps. Both matter, but neither can be attributed or quantified reliably. Leaving them unmodeled protects the credibility of the eight supported lines.

The revenue math, in detail

Pipeline Velocity = deals/qtr x avg deal value x 4 x conversion lift. Two of the three inputs, deals per quarter and average deal value, come straight from the customer's CRM. The challenging input is conversion lift: the claim that handling leads faster and more accurately actually closes more of them. The strongest source is the customer's own data; failing that, a published speed-to-lead benchmark, checked against their funnel, stands in.

Pricing this conservatively, a 3.5% lift on 1,000 deals a quarter at \$60K each, puts Pipeline Velocity near \$8.40M a year:

Conversion lift	Basis	Annual value
2.0% (floor)	1,000 x \$60K x 4 x 2.0%	\$4.80M
3.5% (used here)	1,000 x \$60K x 4 x 3.5%	\$8.40M
5.0% (upper)	1,000 x \$60K x 4 x 5.0%	\$12.00M

The combined case

Every line below traces to a formula and a sourced input the customer can inspect:

Value driver	Dimension	Illustrative inputs	Tier	Annual value
Pipeline Velocity	Revenue Impact	1,000 deals/qtr, \$60K, 3.5% lift	B	\$8.40M
Revenue Capture	Revenue Impact	\$600M revenue, 1.0% leakage, 33% capture	B	\$2.00M
Time to Revenue	Revenue Impact	750 new/yr, \$60K, 15 days	B	\$1.85M
Task Elimination	Productivity	15,000 tasks/mo, 8 min, \$55/hr	A/B	\$1.32M
Labor Avoidance	Cost Avoidance	15 FTEs avoided, \$130K loaded	A	\$1.95M
Error & Rework Elimination	Cost Avoidance	2,500 errors/mo, \$60, 70% cut	B	\$1.13M
Tool Consolidation	Cost Avoidance	15 tools, \$90K/tool/yr	A	\$1.35M
Data Integrity	Risk & Quality	50,000 records/mo, 5%, \$50, 60% cut	C	\$1.80M
Incident Prevention	Risk & Quality	10 incidents/yr, \$200K avg, 45% cut	C (directional)	\$0.90M
Total				\$20.70M

It helps that several lines carry no soft assumption at all. Labor Avoidance and Tool Consolidation are built entirely from counts the customer can verify, a Tier-A floor of about \$3.3M, and Handoff Elimination is a pure elapsed-time result booked under Speed. Even if a buyer set aside the entire revenue block, about \$8.45M of the case would still stand, so the total doesn't hinge on the conversion-lift argument the way a revenue-heavy case would. A strong revenue number is still only half the decision, though — this workflow gets prioritized because it also clears the feasibility bar covered below.

By dimension, Revenue Impact is about \$12.25M, roughly three fifths of the total. Cost Avoidance, Risk & Quality, and Productivity split the rest, at \$4.43M, \$2.70M, and \$1.32M. Speed stays a cycle-time result with no dollar figure attached.

That \$20.70M is gross value, before any feasibility discount. Discounted by 25% to reflect feasibility, it leaves about \$15.53M; subtracting a \$1.70M all-in cost to build and run the workflows lands a realizable value of roughly \$13.83M.

Everything above is the top-down story: specific use cases, each mapped to a driver with a formula behind it. It's the more defensible half, and it's what this example builds, but it's not the whole picture. A second source of value shows up once a platform is in broad use, when people across the business build their own automations rather than waiting on a central team.

On illustrative inputs, a company of 5,000 with about 300 of them building at two hours saved a week, valued at \$100 an hour across a working year, is worth about \$3M a year on its own, a meaningful share of the value a top-down count alone would miss.

That bottom-up number is harder to source and should be presented as directional rather than booked alongside the driver totals, but it belongs in the case because leaving it out understates the potential return.

Answering the questions finance will raise

Four questions come up in nearly every review, and a good model has already answered each one:

How will we know the automation caused it? Capture the baseline before go-live so a before-and-after exists; where that's not possible, use a hold-out group or a cited benchmark.

Are you double counting? No. Each saved hour and each saved dollar sits under exactly one driver. Speed measures elapsed calendar time, so it never carries a dollar figure; the value of the effort that time frees up is booked once, under Productivity. The faster response time above is a Speed result for that reason, and its dollars appear only under Productivity, never in both.

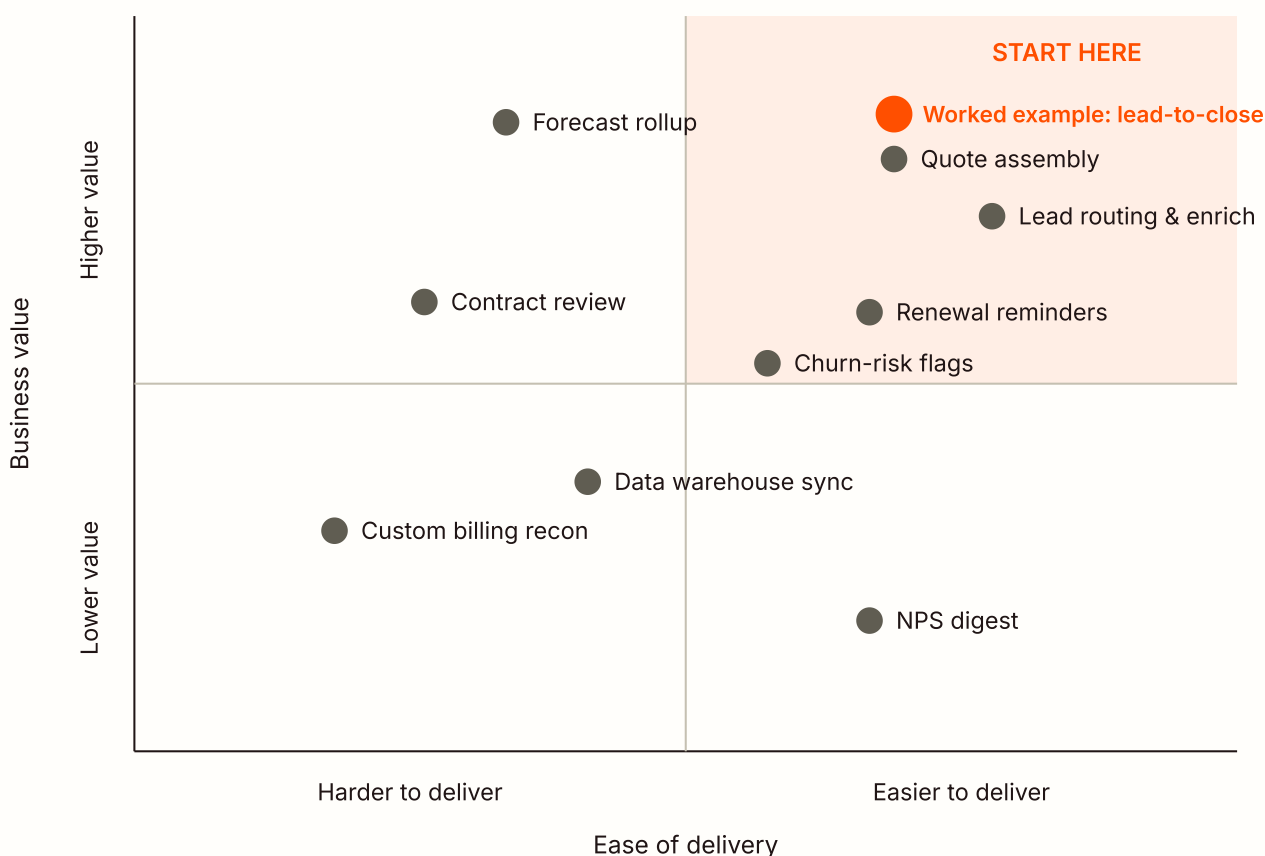
Where did the benchmark come from? Every default carries a citation and an evidence tier (Appendix B); anything without a source doesn't enter the model.

Will it even ship? This is what the value model alone can't answer, and it's what the feasibility score below is for.

Prioritizing the portfolio

Feasibility is scored across five factors: integration complexity, workflow structure, data readiness, stakeholder scope, and time to live. The full rubric is in Appendix C. The overall rating is the worst of the five, not the average: perfectly clean data still rates High overall if the project needs a security review, because the review sets the timeline.

Plot every candidate on the same value-and-effort grid from page one. The low-effort, high-value use cases go first; once they're returning, it's easier to build momentum for the cross-functional projects that need a security review and engineering.



Illustrative portfolio. Fund high-value, low-effort use cases first.

Figure 3. The same grid with real candidates plotted. Illustrative portfolio.

What to do next

Want to know how to build your automation strategy? Our business teams and solutions architects will [run this assessment](#) with you using the framework laid out here. You come away with a ranked set of projects, each sized for value and scored for feasibility, with a formula and a source behind every figure, and an answer to which of the worthwhile projects to build first.

Everything past this point is reference material: the full driver formulas, evidence-tier rules, feasibility rubric, the build process, a sample-project menu, and a glossary. Pull it up when a specific number needs defending.

Appendix A: The full framework – five dimensions, seventeen value drivers

Every driver comes down to a formula: a few customer counts and one or two impact assumptions. The default benchmark values behind them live in the ROI model.

Dimension	Value driver	Formula
Revenue Impact	Pipeline Velocity	deals/qtr x avg deal value x conversion lift x 4
Revenue Impact	Revenue Capture	annual revenue x leakage rate x capture improvement
Revenue Impact	Revenue Expansion	customer base x expansion rate x avg expansion value x lift
Revenue Impact	Time to Revenue	new customers/yr x revenue/customer x (days accelerated / 365)
Speed	Process Acceleration	processes/mo x (time before – time after) x 12 [elapsed time; dollar value of freed effort is booked under Productivity]
Speed	Handoff Elimination	handoffs/mo x avg queue time x 12 [elapsed time; dollar value of freed effort is booked under Productivity]
Productivity	Task Elimination	tasks/mo x minutes/task x (rate / 60) x 12
Productivity	Task Simplification	tasks/mo x minutes saved/task x (rate / 60) x 12
Productivity	Context Surfacing	meetings + searches avoided x time x rate x 12
Productivity	Ramp Acceleration	new hires/yr x monthly target x months accelerated x productivity gap
Cost Avoidance	Labor Avoidance	FTEs avoided x fully-loaded annual cost
Cost Avoidance	Tool Consolidation	tools eliminated x annual license cost per tool
Cost Avoidance	Error & Rework Elimination	errors/mo x cost per error x reduction rate x 12
Risk & Quality	Compliance Assurance	expected violations/yr x avg penalty x reduction rate
Risk & Quality	Data Integrity	records/mo x error rate x cost per error x reduction rate x 12
Risk & Quality	Incident Prevention	incidents/yr x avg cost per incident x reduction rate
Risk & Quality	Process Consistency	processes/mo x defect rate x cost per defect x reduction rate x 12

Appendix B: Evidence tiers

In principle, everything can be quantified. In practice, many numbers will be indeterminate or contested. It takes judgment to know what not to quantify, so that the weak numbers don't drag down the strong ones. That is what the evidence tiers are for. Every number carries a tier that notes where it came from, so a buyer can weigh it rather than take it on faith, and the structure moves the conversation away from whether to believe a single total and toward which specific inputs to adjust, a conversation a finance team or line-of-business team can actually engage with.

There are three tiers. Tier A is the customer's own figure, used as is, and it is the strongest input available. Tier B is a published study, used with context and checked against the customer. Tier C is a case study or other industry data, treated as directional and flagged as single-source. When in doubt, tier down. A Tier B you can't actually cite is really a Tier C. Leave out any number too weak to stand beside the credible lines. The value may be real, but a shaky figure undermines the whole model.

Not every input carries the same risk. Within a driver, the customer's own counts, deal size, headcount, volumes, are low risk. The number that needs defending is the impact assumption: the conversion lift, the reduction rate, the percentage of time saved. A pre-filled figure there isn't a neutral default; it is a claim made on Zapier's behalf.

The rule follows from that: an impact assumption is either the customer's own number, which makes it Tier A, or an explicitly sourced benchmark. It is never an unsourced estimate.

Appendix C: The feasibility rubric

Value only measures the size of the potential impact; it says nothing about whether you can reach it. Feasibility is scored on five factors, which together supply the two things the value equation cannot: the probability of capture and the build cost. Of the five, data readiness is the factor most likely to turn a promising project into a hard one — CRM data that looks clean at the start frequently needs more work than expected, so it's worth pressure-testing early rather than late.

One clarification: this paper uses the word "risk" in two different senses. Implementation risk (used above) is anything that lowers the odds of capture — a brittle integration, an unowned data source, a security review with no end date. The Risk & Quality *dimension* in the value model is a separate concept, measuring the cost of the bad outcomes a workflow prevents.

Criterion	Low	Medium	High
Integration complexity	Apps in the stack, native triggers	1 to 2 new connections or a webhook	Custom API and auth, no native integration
Workflow structure	Linear, 5 steps or fewer	Branching, 5 to 15 steps	Multi-path with error handling, 15+ steps
Data readiness	Clean and structured	Some mapping needed	Quality issues, multiple sources
Stakeholder scope	Single owner	2 to 3 teams or admin sign-off	Security review, change management
Time to live	Days	1 to 3 weeks	4 to 8 weeks

Appendix D: How the business case gets built

All of this comes out of a fixed sequence that starts with the customer's experience and their own operating figures, and ends with a value case built with them.

Step	What happens
Pull account data	Combine Zapier usage, such as task counts, apps, and active users, with company research on size, industry, and revenue. This only applies if the company has an existing automation footprint we can observe.
Map to dimensions	Read the apps and task patterns to see which of the five dimensions the account's work actually touches.
Identify use cases	Map each specific workflow to a single value driver, such as lead routing to Pipeline Velocity.
Populate formulas	Fill in the customer's own counts, apply sourced benchmark defaults where they exist, and flag every soft input for validation.
Tag evidence tier	Attach an A, B, or C tier to each number so the customer sees the strength behind it.
Build the output	Assemble the headline value, the dimension breakdown, the use case detail, and the methodology.
Present and validate	Walk it through as a partnership. This is our framework and our math. Where the customer has a better number, we use theirs.

The last step is a working session with the customer. A finished number invites a yes-or-no verdict and little else. Exposing the inputs and letting the customer replace the ones they disagree with turns the case into something they helped build, and in our experience a case the customer helped build is both more credible and easier for them to defend internally.

Appendix E: Sample projects by function

A starting menu of workflows by business function, each mapped to the value dimension it moves first. Treat it as a prompt for the assessment conversation, not an exhaustive list.

Function	Automation	Primary dimension
Sales	Inbound lead routing and enrichment	Revenue Impact
Sales	Quote and proposal assembly	Speed
RevOps	Pipeline hygiene and deduplication	Risk & Quality
RevOps	Forecast rollup and variance flags	Productivity
MarkOps	Campaign lead sync and scoring	Revenue Impact
MarkOps	Content brief intake and triage	Productivity
Support	Ticket triage and routing	Speed
Support	Escalation and SLA monitoring	Risk & Quality
Finance	Invoice intake and GL coding	Cost Avoidance
Finance	Renewal and dunning reminders	Revenue Impact
People Ops	New-hire provisioning and onboarding	Productivity
IT	Access review and audit logging	Risk & Quality

Appendix F: Key terms

Term	Definition
Realizable value	Gross value discounted by the probability you capture it and the cost to build and run it. The number worth funding.
Feasibility check	A five-factor read on how hard a workflow is to deliver: integration, structure, data, stakeholders, and time to live.
Hybrid workflow (fluid determinism)	A workflow where AI handles the judgment step and fixed, repeatable steps handle everything else, so the output is measurable. Zapier refers to this approach as fluid determinism.
Value driver	A single, formula-based way automation creates value, such as Pipeline Velocity. Seventeen drivers sit under five dimensions.
Evidence tier	A source label, A, B, or C, that states how strong the number behind an input is.
Durable workflow	A workflow that keeps running month over month in production, not a one-time test.

Disclaimer: All figures in this paper are illustrative and sized to a representative enterprise. In a real assessment, every input is either the customer's own number or a sourced benchmark, tagged with its evidence tier.