CHAPTER 2

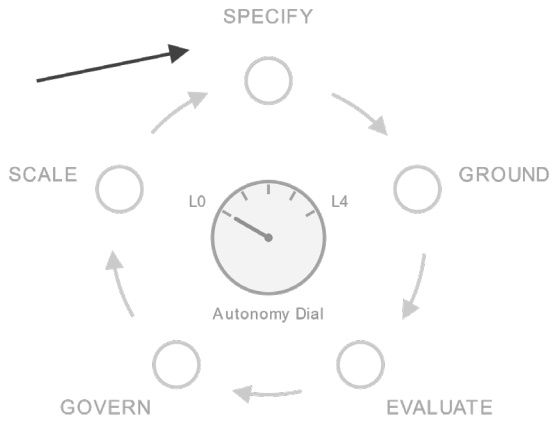
Picking Your Battles

02

Workflow Selection Is the First AI Decision

SKIP PATH

If you already run a scored portfolio of candidate AI workflows — a stack-ranked winner, an explicit posture decision, and a parking lot your executives respect — skip to Chapter 3 (Model Choice Is a Posture, Not a Pick). If your AI roadmap is a wishlist, a mandate, or a single sentence that begins with "we should," read on: this chapter is where AI strategy stops being a slogan.



This chapter sits at the mouth of the Loop. Workflow selection decides what enters the Loop, if at all — nothing is highlighted yet, because nothing has earned it.

Tom Brandt, Meridian's CEO, returns from an industry conference with a video on his phone: a competitor's AI demo. It shows an employee-facing benefits assistant answering a question about dependent coverage — fluent, warm, instant. He has watched it more times than he will admit. By Wednesday, "AI strategy" appears on the leadership agenda. By Friday it has sharpened into the sentence Dana Okon, his VP of Product, will be living with for the next three quarters: "I want AI everywhere by Q3."

Dana does not argue. Arguing with a mandate is how product leaders lose the only meeting that matters. Tom is not wrong that the market is moving; he is wrong about what moving looks like, and the difference between those two errors is the difference between a program and a pileup. The answer to "AI everywhere" is not "no." It is "AI, but in order." Dana asks for one meeting, one wall, and every AI idea anyone at Meridian has ever floated.

You will be in a version of this meeting soon. The most consequential AI decision this year will not involve a model name, a vendor contract, or a prompt. It is which workflow you point AI at, and at what posture. Get it right, and everything downstream gets easier: the behavioral contract has a defined job to specify, evaluation has something concrete to measure, and the economics can be tested. Get it wrong, and no model on earth can save you, because you've aimed a probabilistic instrument at a job that punishes probability.

This chapter gives you the Workflow Selection Scorecard, and something rarer: permission. Permission to say that a workflow everyone loves is not ready. Permission to park it without killing it. Permission to tell the room that the demo was someone else's roadmap.

SECTION 2.1

The Anatomy of a Workflow

Before you can select workflows, you have to see them clearly, and most AI roadmaps never do. They list outcomes ("reduce support costs"), technologies ("add a chatbot"), or aspirations ("transform the employee experience"). None of those is a workflow. A workflow, for selection purposes, has three parts: inputs, outputs, and a failure cost. Everything the Scorecard measures traces back to this anatomy.

Inputs are what arrives, in whose language, and in what condition. A support ticket is an input: messy, human, ambiguous, written by someone who doesn't know your product's vocabulary. A payroll file is also an input: structured, exact, and intolerant of interpretation. The first tells you a language engine might belong here. The second tells you it might not.

Outputs are what the system must produce — and, just as important, who consumes them. When a trained colleague reviews an output before anyone acts on it, a wrong answer can be caught, corrected, and learned from. An output sent straight to a customer, an employee choosing a health plan, or another system that acts on it immediately has bypassed that review. The audience of an output is the single best predictor of how much control the workflow will demand.

Failure cost is what a wrong output costs, and where the cost lands. Meridian's ticket triage has a modest one: a misrouted ticket wastes a rep's minutes, and a flawed draft response can be edited before it ships, because a human sends every reply. The employee-facing benefits advisor in Tom's video has a different anatomy entirely: a wrong answer about enrollment windows or COBRA deadlines lands in someone's life, not in a queue, and often surfaces months later — when the harm is done and the trail is cold. Same company, same underlying capability, radically different exposure. Failure cost is a property of the workflow, not the model.

Multiply by volume (how often the workflow runs) and you have the selection picture. Workflows whose anatomy you cannot describe this concretely are not candidates yet. They are conversations.

SECTION 2.2

The Five Disqualifiers

Before you rank candidates, remove the ones that should not enter the ring as proposed. A large language model is the right tool when language ambiguity is the bottleneck and you can tolerate probabilistic behavior with controls. That sentence has five ways to be false, and each one is a disqualifier. Any single one is enough.

One: the task demands determinism. If the same input must produce the identical output every time (billing calculations, entitlement logic, access control, exact policy enforcement), you need conventional software, because a language engine is a probability instrument and no prompt makes it otherwise. The productive pattern is division of labor: let the model interpret the messy human input at the edge, then hand the decision to deterministic code. Meridian's enrollment

eligibility rules will surface in the meeting below wearing an AI costume they don't need.

Two: a wrong answer is unacceptable, and you cannot add controls. High-stakes domains (medical guidance, legal conclusions, financial approvals, benefits determinations) can be served by AI only inside a cage of grounding, strict refusal rules, logged evidence, and human review. If the workflow cannot accommodate that cage, because the channel is real-time or the volume swamps reviewers or the user is a stranger, do not deploy. Meridian has already run this experiment. The 2025 beta granted a caller a 90-day COBRA grace period that did not exist — *Confident Nonsense*, delivered with perfect fluency into the highest-stakes conversation the company has. Whatever the model's capability, Meridian had pointed it at a workflow with no tolerance and no cage.

Three: no authoritative source of truth exists — and you can't create one. If your organization's knowledge is scattered, inconsistent, or politically contested ("the policy depends on who you ask"), the model will reflect that mess. AI will not fix your governance; it will amplify your confusion in confident prose. When this disqualifier fires, the AI project hasn't started yet: you have a knowledge governance project, and Chapter 7 is about what happens when you skip it.

Four: the workflow can't tolerate latency, variability, or limited explainability. Model calls cost money and time. Outputs vary between runs. And even with citations, you won't always have a tidy account of why a particular phrasing appeared. Sometimes variance is a feature — drafting, ideation, exploration. Sometimes it's a defect — compliance language, customer commitments, anything audited word by word. If the user experience requires instant, identical, fully explicable output, simpler machinery wins.

Five: you're automating a process you haven't standardized. If the humans who run the workflow cannot describe its decision criteria consistently, a model will not operationalize your ambiguity. It will guess, differently each time, and the inconsistency will read as malfunction. Automate a messy workflow and you get automated mess. Standardize first — which is human work, not model work — or choose a posture humble enough to survive the mess.

Notice what disqualification means. It doesn't mean the idea is worthless. It usually means the idea is a different kind of project — deterministic software, knowledge governance, process standardization — and discovering that before you spend is the cheapest insight in this book.

SECTION 2.3

The Four Questions

The five disqualifiers compress into a gate you can run in any executive room: four questions, each answerable yes or no, asked of every candidate workflow before any scoring begins.

1. **Is the core input mostly language?** Messy text that needs structured meaning — not arithmetic wearing a costume.

2. **Can we ground answers in sources we own and trust?** There is an authoritative truth, or a credible plan to build one.

3. **Can we measure and monitor the behavior?** Success is definable and checkable, before launch and forever after.

4. **Can we contain the risk with workflow design?** Human-in-the-loop (HITL) review, permissions, refusal rules, escalation paths — controls proportional to the failure cost.

Any "no" parks the workflow, whatever the upside. The gate is the Operating Loop interviewed in advance: the second question is **Ground** asked before the work begins, the third is **Evaluate**, the fourth is **Govern**, and the first asks whether a language engine belongs in the room at all. A workflow that cannot answer these questions cannot survive the Loop.

The first question asks about the input, not the output. A system that returns an image, a clip, or a synthesized voice is still specified in language — the prompt is the contract either way. What the output modality changes is how you check the result: exact matching gives way to a rubric and a sample, which is Chapter 5's problem, not this gate's.

The gate also protects you from the strongest force in AI adoption, which is not technology. Don't do AI because a competitor did. "Everyone has an AI assistant" is not a product requirement. It's peer pressure. Tom's conference video is a requirement on someone else's roadmap: evidence about a competitor's appetite, not Meridian's readiness. The gate converts that pressure into questions with answers.

IN THE ROOM

The questions to carry into any AI proposal, and the scripts for the hard moments:

When someone brings a mandate: "Which workflow first? We'll get to 'everywhere' faster through one battle won than twelve declared."

When someone brings a demo: "What's the failure cost, and where does it land — in a queue we control, or in a customer's life?"

When someone claims the data is ready: "Where does the truth live, and does it agree with itself?"

The discipline question, asked of every candidate: "What would make us park this?" If the honest answer is "nothing," you are not selecting. You are complying.

The redirect that isn't a veto: "We'll reach that workflow faster by not starting with it. Here's what has to be true first — and here's what we start with while we make it true."

SECTION 2.4

The Workflow Selection Scorecard

With the gate in place, you can rank what survives it. The Scorecard is a two-stage instrument: the gate decides *whether* a workflow may enter the portfolio; the score decides *when*. Four dimensions, each scored 1 to 5, summed to twenty.

Value (volume × pain × leverage). How often does this run, how much does it hurt, and how much does relief compound? A 1 is an occasional annoyance nobody has costed. A 5 is a daily, high-volume workflow with a cost line an executive already resents.

Tractability. Is language ambiguity the actual bottleneck, and can you ground the work in sources you own? A 1 is deterministic logic in

a language costume. A 5 is a job whose whole difficulty is reading, classifying, or drafting messy language, with the evidence to ground it sitting in systems you control.

Failure Tolerance. What does a wrong answer cost, and can workflow design contain it? A 1 is irreversible harm landing outside your building. A 5 is a wrong answer that is cheap, visible, and caught by a human before it matters — containment by design, not by luck.

Data Readiness. Does an authoritative source of truth exist? A 1 is knowledge that is scattered, contested, or stale. A 5 is a single owned, current, versioned source that the organization already treats as canonical.

Weight the dimensions equally by default. If this is your organization's first AI workflow, double-weight Failure Tolerance: your first workflow's real product is a learning loop, and the program's survival is worth more than any single workflow's upside. Once the discipline in this book is running, shift weight toward Value: you've earned the right to chase upside. Never weight any dimension to zero; a zero weight is a disqualifier you've decided not to notice.

Two rules give the instrument its teeth. First, the disqualifier rule: **any "no" on the four questions parks the workflow regardless of score.** A nineteen that fails the gate is parked next to the elevens. Score ranks the eligible; it never overrides the gate. Second, every workflow receives exactly one of four verdicts: **Build** (one winner, two at most), **Queue** (eligible, sequenced, named blocker), **Park** (fails the gate; recorded reason, re-review trigger, owner), or **Disqualify** (not an AI project — route it to software or governance where it belongs).

Do not mistake the Scorecard for a measurement of truth. The numbers are coarse on purpose; their job is to force the argument. Score fast, in a room, with the people who run the workflows. And when two people score a row differently, the disagreement is the data. The Scorecard's product is not a ranking. It is a defensible no.

SECTION 2.5

The Selection Meeting

Dana spends a week collecting candidates: from support, from sales, from HR operations, from marketing, and from Tom's phone. Twelve survive a first sanity pass and go on the wall. In the room: Marcus Webb, who runs support operations and sits with his arms crossed, because he has been burned by a chatbot pilot before and the Phantom Policy is still fresh; Priya Raman, who must build whatever wins; Elaine Cho, Meridian's general counsel, who must defend it; and Tom, who wants all twelve.

Marcus supplies the ground truth nobody else has: 52,000 tickets a quarter, $8.40 fully loaded per resolved ticket, a 3.2-hour median first response, 14 percent of tickets escalated. Benefits-policy questions make up roughly 31 percent of that volume, and during open enrollment weekly volume runs about two and a half times normal. The wall gets scored in ninety minutes. It looks like this:

#	Candidate workflow	V	T	F	D	/20	Four-question gate	Verdict
1	Triage – support ticket triage + drafted responses for rep review	5	5	4	4	18	Yes ×4	**Build** — co-pilot
2	Support-facing benefits-policy Q&A	4	5	4	2	15	Yes; Q2 conditional	Queue — behind source-of-truth work
3	Employee-facing benefits advisor	5	4	1	1	11	**No: Q2, Q4**	**Park** — triggers set (see below)
4	Refund & credit recommendations	3	3	2	4	12	Yes; Q4 needs controls	Queue — after containment is proven
5	COBRA & compliance correspondence drafting	4	4	1	3	12	**No: Q4**	Park — review can't keep pace with risk

#	Candidate workflow	V	T	F	D	/20	Four-question gate	Verdict
6	Open-enrollment plan-configuration checks	4	1	2	4	11	**No: Q1**	Disqualify — deterministic software
7	Account-team call summarization	2	4	5	3	14	Yes ×4	Queue — low value, low urgency
8	RFP & security-questionnaire drafting	3	4	3	2	12	**No: Q2**	Park — product facts scattered
9	Carrier file-error triage	3	2	3	4	12	**No: Q1 (mostly)**	Disqualify as scoped — rescope the language layer
10	Leave-eligibility determination	4	1	1	3	9	**No: Q1, Q4**	Disqualify — entitlement logic, zero tolerance
11	Marketing content generation	1	4	4	2	11	Yes ×4	Queue — bottom; see below
12	KB article drafting from resolved tickets	3	4	4	3	14	Yes ×4	Queue — companion to knowledge cleanup

The gate does its work before the scores say a word. Row 6 is Priya's favorite verdict of the afternoon: "That's a fine project. It's just software's job — the rules are already exact, and exact is the one thing we'd be paying a language model to ruin." Row 10 goes the same way, harder: eligibility determinations are entitlement logic with legal consequences, a "no" on two questions at once. Row 9 gets rescoped on the spot: the file errors are structured codes, and only the explanation layer is language-shaped. Row 11 goes to the bottom of the queue in the shortest conversation of the meeting: nobody can name a pressing problem this workflow would solve.

Row 3, the employee-facing benefits advisor in Tom's video, ties the winner for the highest Value score. It is parked, because value was never the question. Elaine takes failure cost: a wrong answer about an enrollment window or a COBRA deadline is not a support error, it is somebody's uncovered surgery, and it surfaces months after the conversation that caused it — the one failure mode Meridian has already field-tested, on a recorded line. Marcus takes truth: the knowledge is spread across three systems that disagree, and his reps arbitrate those disagreements all day. Two questions, two honest nos.

Tom pushes, as he should. He knows what the competitor is shipping. Dana agrees with him. "It's the best idea on the wall, and we can't operate it yet. AI is not a feature. AI is a system. The system this one needs — one source of truth, containment we've proven on cheaper failures — doesn't exist here yet. Triage is how we build it." The employee-facing benefits advisor is parked, not killed, with its re-review triggers written down: the knowledge base consolidated to one owned source of truth, and containment proven on lower-stakes workflows.

Row 1 wins on every dimension that matters: 52,000 runs a quarter of messy human language, evidence inside Meridian's own systems, and a failure cost contained by a control already built into the workflow — a human reviews and sends every reply. Row 2 is the clear second, gated on the knowledge cleanup it will eventually force. The meeting ends with one battle chosen, not twelve declared.

SECTION 2.6

The Parking Lot Is a Strategy

What Dana built in that meeting is not a roadmap with some rejections attached. It is a portfolio — and the parking lot is the part that makes it one.

Parking is a strategic act, not a defeat, and the distinction lives in the paperwork. Every parked workflow leaves with three things: its reason, stated in gate terms ("failure cost uncontainable at current review capacity," not just "not ready"); a re-review trigger — the

event that reopens the case, such as a consolidated source of truth, a proven containment pattern, or a capability shift worth re-testing; and an owner who will recognize the trigger. Parked means not yet. A parking lot without triggers is just a graveyard with better signage.

The portfolio needs kill criteria too, because optimism accumulates. A queued workflow whose blocker misses two consecutive re-reviews gets killed or re-scoped. A parked workflow whose value case evaporates — the process changed, the volume moved — comes off the lot entirely. Killing a candidate in a portfolio review costs a sentence; killing it after two quarters of building costs a team. Meridian re-scores the portfolio monthly. What you learn operating the winner re-scores every candidate behind it. Chapter 13 shows that cadence as part of the full governance rhythm. Selection is never finished. It is only currently passing.

> **FROM THE FIELD**
>
> McDonald's spent roughly three years piloting voice-AI order-taking in drive-thru lanes with IBM before ending the program in 2024, after viral videos of confidently wrong orders — bacon on ice cream, an order of 260 unwanted McNuggets — exposed errors to a public audience. Read it as a selection failure, not a model failure. The input was unbounded speech in a noisy channel; the output was customer-facing, so ordering mistakes reached customers directly; and the posture was automation before the workflow had earned it. The same capability, pointed at kitchen-facing co-pilot work, could have made failures cheaper and less public.

SECTION 2.7

The First Rung Question

One decision remains, and it's the one most selection processes skip: not just *which* workflow, but *at what posture*. Posture is the co-pilot-versus-automation choice: how much the system does with a human, versus instead of one. You make it explicitly, or discover you made it implicitly during an incident.

Co-pilot means AI assists a human who owns the decision. Adoption comes faster: nobody's job disappears into a black box on day one.

Errors are cheap when a person catches them in the flow of work. Ambiguity is survivable; human judgment absorbs what the model fumbles. The return has a ceiling, though: if humans stay the bottleneck, executives eventually ask why the expensive assistant still needs an escort.

Automation means AI executes with humans handling exceptions. The return scales with volume instead of headcount, and the price is steep: the workflow must be standardized, the behavior measurable, the exception paths designed and tested, because failures are no longer private edits in a queue. They are outputs in the wild, visible and compounding.

The progression that works is the one the QuickStart named: co-pilot first, automation later — unless the workflow is already standardized and measurable. That clause is the first rung question of Chapter 10's Autonomy Ladder. It runs from L0, where AI merely suggests, through L4, where it acts toward goals inside hard boundaries, to L5, where agents run as a workforce. This book names L5 but leaves it unpriced; its gates are engineering, not product. Co-pilot lives on the bottom rungs (L0 and L1), where a human performs or approves every action. Automation lives above, and every rung up is bought with proof. The Ladder's first law is already a selection criterion: autonomy is earned, not configured. Every workflow starts at the bottom by default. The posture question is not "co-pilot or automation, forever?" It is: *where does this workflow start, and how high must it climb for the business case to hold?* If the business case closes only at a rung the workflow cannot yet earn, that is a selection fact — and the parking lot is where it belongs. That is precisely what happened to the benefits advisor: Meridian cannot yet ground its answers or contain its errors.

Meridian sets triage at co-pilot, L0 to L1: the model routes tickets and drafts responses, and a rep approves everything that leaves the building. It's Marcus's condition for cooperation, and it's the right call for reasons that have nothing to do with his scar tissue — the workflow is ambiguous, the trust is unbuilt, and human review is already part of the workflow.

If Meridian were this book's only evidence, you could mistake co-pilot for a rule rather than a verdict. So meet the counterexample: Castellan,

a commercial-lending firm that extracts terms from master service agreements — 1,400 a quarter, the same clauses, the same fields, every time. Their inputs are standardized documents. Their outputs are machine-checked: every extracted term points back to a clause, and validation is code, not judgment, catching errors before money moves. Castellan chooses automation from day one, and they're right — the same way Meridian is right to choose the opposite. Same gate, same scorecard, opposite answer, because posture is a property of the workflow, not a measure of nerve.

SECTION 2.8

What Dana Sends Tom

The meeting's output fits on one page. The winner: ticket triage with drafted responses, co-pilot posture, L0 to L1.

The queue: policy Q&A behind the knowledge cleanup, then the rest in order, each with its named blocker.

The parking lot: the advisor, the compliance drafting, the RFP responses — each with its reason, trigger, and owner. The rerouted: two "AI ideas" sent to the software backlog where they always belonged. The closing line for portfolio review: what we learn on triage re-scores everything below it.

Tom reads it, circles the benefits advisor's re-review triggers, and writes "Q3?" in the margin. He is still Tom. But he funds the page — because it's the first version of "AI everywhere" that comes with an order of battle instead of a wish. The decision must also survive audit and reversal. The next chapter introduces the formal instrument for that, the Decision Record, and this page becomes Meridian's first entry in the file.

MONDAY MORNING

1. List every AI workflow your organization has promised, piloted, or wished for — that list is your portfolio, whether or not you manage it as one.

2. Run each candidate through the four questions, and park every "no" with a written reason and a re-review trigger.

3. Score the survivors on Value, Tractability, Failure Tolerance, and Data Readiness; stack-rank them and pick one Build.

4. Set the winner's posture — co-pilot unless the workflow is already standardized and measurable — and write down what would justify climbing.

5. Publish the whole portfolio, parking lot included, to the executives who asked for "AI everywhere."

Dana has a workflow and a posture; what she doesn't have is an engine — and Chapter 3 is about why choosing one is not the decision you think it is.