

AGENTIC ENGINEERING

FROM FIRST PRINCIPLES

BUILDING, CONTROLLING, AND SCALING
LLM AGENTS IN PRACTICE

From a Single Model Call to Autonomous Systems
That Use Tools, Learn, and Improve Themselves



CHRISTIAAN VAN DER WALT, PhD

Agentic Engineering for Executives

The questions to ask before you approve an agent

Central question

Where should your organisation let software act on a model's decisions, how much autonomy should it grant, and who answers for the result?

A briefing for leaders who will approve, fund, and answer for agentic systems

The short version

An agent is software that lets a model's decisions cause actions. Grant it the least autonomy that achieves the result, start where outcomes can be checked and actions undone, measure cost per completed task, put a named person in charge of every agent, and expand only on evidence.

The rest of this book is written for the people who build agents. This briefing is written for the people who decide whether they should be built, where, and under what control. It assumes no programming knowledge and takes about forty minutes to read.

Executive books on artificial intelligence explain well why AI matters and how organisations should change around it. They say less about the engineering facts that decide whether an agent project pays off: when an agent is better than a simpler workflow, what really drives its cost, which failures to expect, and what “under control” means in practice. Those facts are what this book measures. Each recommendation below is tied to the chapter, and usually to the experiment, that demonstrates it, so you can hand the chapter reference to your technical team and ask them to show you.

Figures from outside the book are dated and sourced, because this field moves quickly. Treat them as the state of evidence at that date, not as permanent truths.

1. The decisions that are yours

Most decisions about agents belong with engineers. A small number cannot be delegated, because they set the level of risk the organisation accepts.

Decision	Why it sits with leadership	Section
Where to use agents	It commits budget and people, and it decides which processes change.	4
How much autonomy	It sets how much can go wrong before a person notices.	3
What it may cost	Agent costs grow with every step, retry, and escalation.	5
Who is accountable	An agent cannot be held responsible; a person must be.	7
What counts as success	Without agreed measures, demos substitute for evidence.	8
Buy, build, or partner	It decides whether the capability stays in the organisation.	10
When to stop	Many projects fail because nobody defined the exit.	11

2. What an agent is, in one page

A language model on its own produces text. It can describe an action, but it cannot take one. An *agent* is a software system in which a model's decisions can cause actions, and the results of those actions come back to the model as new information for its next decision. That loop of decision, action, and observation is the whole difference (Chapter 1).

Every agent, however it is marketed, is made of five parts:

It helps to place agents next to the systems you already know:

Part	What it is	The executive question
Model	The language model that proposes the next step.	Which model, from whom, and at what cost per call?
Context	The information the model sees when it decides.	Is the information current, permitted, and sufficient?
Actions	The tools and systems the agent can change.	What exactly can it change, and can that be undone?
State	What it remembers between steps and between tasks.	What does it keep, for how long, and who can see it?
Control	The rules that limit, check, and stop it.	Who set the limits, and what happens when they are reached?

System	Who decides the next step
Rule-based automation (RPA)	A programmer, in advance, for every case.
Workflow with an AI step	A programmer fixes the path; a model handles one step, such as reading a document.
Chatbot or copilot	A person, after reading the model's answer.
Agent	The model, at run time, within limits that software enforces.

How to spot “agent washing”

Ask one question: does a model's decision cause an action whose result the system then observes and acts on again? If not, it may be useful, but it is not an agent. In June 2025 Gartner estimated that only about 130 of the thousands of vendors marketing agentic products offered genuinely agentic capabilities^[1].

3. The autonomy decision

Autonomy is not a property of the model. It is a decision you make about how much the system may do before a person looks. It helps to think of it as a ladder.

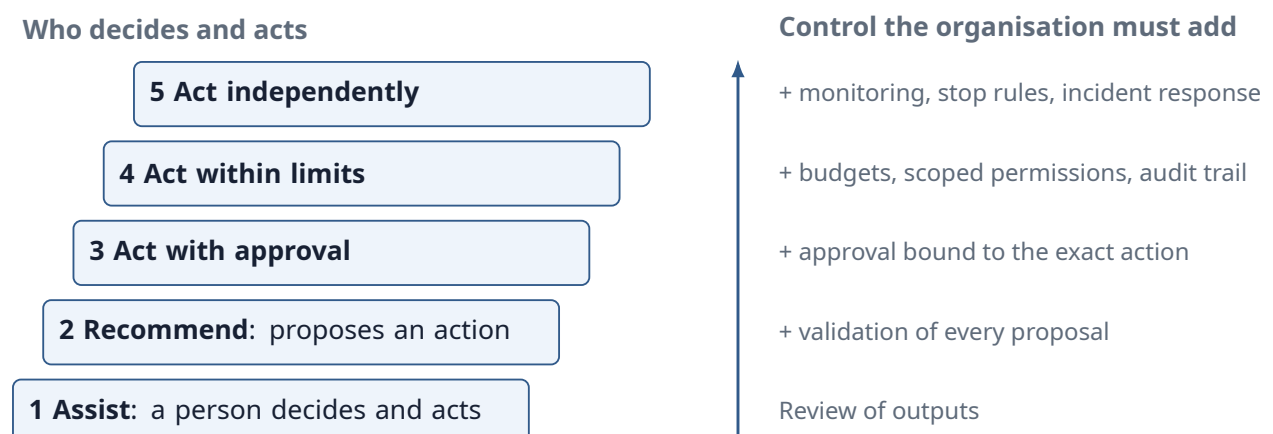


Figure 1: The autonomy ladder. Each step up removes a human check, so each step up must add a software control in its place.

The central rule of this book applies directly to leadership decisions:

Least autonomy

Use the least autonomy necessary to obtain the adaptability the task requires (Chapter 4).

Many tasks do not need an agent at all. If the steps are known in advance, a fixed workflow is cheaper, faster, and easier to audit. Gartner made the same observation when it predicted, in June 2025, that over 40% of agentic AI projects will be cancelled by the end of 2027: many use cases presented as agentic do not require agentic implementations^[1].

Evidence from the book (Chapter 4, Experiment 7)

A report-summarising task is solved twice: once by a fixed workflow, once by an agent. Both produce the same answers. The workflow uses no model calls; the agent uses three for an English report and four for a French one that needs translating. When the path is known, autonomy adds cost and variability without adding value.

Autonomy earns its place where the right next step genuinely depends on what was just found out: investigating an incident, reconciling records that disagree, repairing code until its tests pass. Even then, the step up the ladder should be taken deliberately, one rung at a time, with the matching control in place first.

4. Where the value is

Adoption is not the problem. Results are. In June 2025 McKinsey reported that nearly eight in ten companies it surveyed use generative AI, yet just as many report no significant bottom-line impact, a gap it called the “gen AI paradox”^[2]. In August 2025 MIT’s NANDA initiative reported that about 95% of the enterprise generative AI pilots it studied delivered no measurable profit-and-loss impact, based on 150 interviews, a survey of 350 employees, and an analysis of 300 public deployments^[3]. The 95% figure comes from an early research programme and has been debated, but its explanation is more useful than the number: the tools did not learn from, or fit into, the way work was actually done.

Both reports point to the same lesson. Value comes from redesigning a process around what agents do well, not from attaching an agent to a process designed for people. MIT also found that the largest returns came from back-office work, while more than half of budgets went to sales and marketing tools^[3].

Good first uses share a recognisable profile:

Property	Why it matters
The outcome can be checked	You can tell automatically whether the agent succeeded (Chapter 22). Coding agents thrive because tests can be run (Chapter 28).
Actions can be undone	A mistake becomes a correction, not an incident.
The volume is high	Small savings per task add up; the cost of evaluation is spread wide.
The path varies from case to case	Otherwise a workflow is the better tool (Section 3).
Someone owns the process	A named owner decides what good looks like and answers for it.
The data is reachable and permitted	An agent can only decide well on what it is allowed to see (Chapter 6).

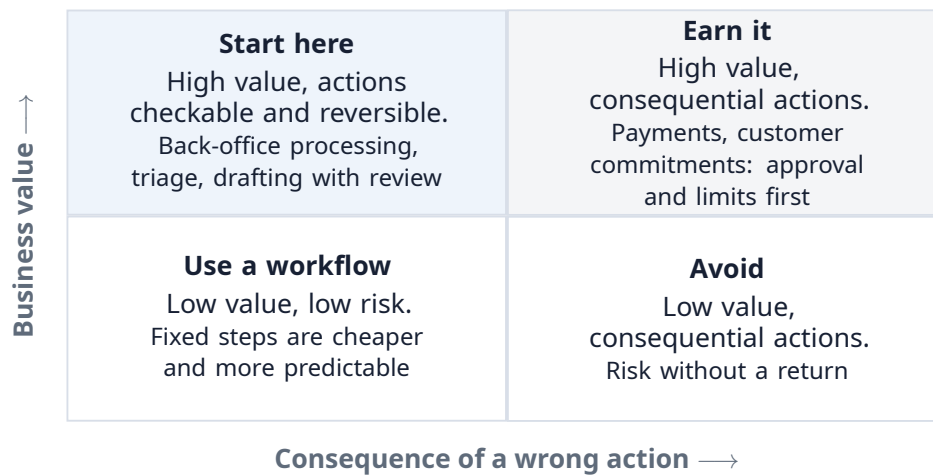


Figure 2: Prioritising a portfolio of agent opportunities by business value and by the consequence of a wrong action.

5. The economics

The price of a single model call is the wrong number to manage. The useful number is the *cost per completed task*, including everything that task required.

- **Steps.** An agent calls the model once per decision. A task that takes eight decisions costs eight calls, and longer contexts make each call dearer (Chapters 3 and 6).
- **Retries and failures.** Failed attempts still cost money. Unbounded retries multiply load (Chapter 24).
- **Escalation.** Routing hard cases to a stronger model raises quality and cost together (Chapter 15).
- **Evaluation and oversight.** Test suites, monitoring, and the time people spend reviewing and approving are part of the cost, not overhead to be ignored.
- **Incidents.** The expected cost of a wrong action belongs in the business case.

Three levers keep the economics sound. Use a workflow wherever the path is known. Route each task to the cheapest model that can be shown to handle it, escalating only when a check says so (Chapter 15). And enforce budgets *before* each action rather than discovering overspend afterwards (Chapter 27).

Evidence from the book (Chapter 27, Experiment 51)

An agent searches for an answer that its sources never contain. With no limits, it makes 100 searches, spends 3.00 cost units, and still has no answer. With a budget checked before every action, it stops after three searches, at 0.09 of a 0.10 limit, and reports why it stopped: BUDGET_EXHAUSTED.

Costs, unclear value, and weak risk controls are exactly the three reasons Gartner gave for its prediction that over 40% of agentic AI projects will be cancelled by the end of 2027^[1]. All three can be addressed before a project is approved.

6. What goes wrong

Agents fail in characteristic ways. None of these failures is exotic; each is reproduced deliberately in this book so that the matching control can be engineered. You do not need the details, but you should expect your teams to have an answer for every row.

Failure	What it looks like	Control (chapter)
Invalid action	The model proposes an action that does not exist or has the wrong inputs.	Validation before execution (2)
Runaway loop	Every step looks fine, but the task never finishes.	Step limits and budgets (3, 27)
Prompt injection	Text in an email or web page instructs the agent to do something else.	Authority kept outside content (26)
Poisoned memory	A false “fact” is stored and trusted later.	Write policies for memory (8, 26)
Agreeing on the same mistake	Several agents or votes confirm a shared error.	Independent evidence and checks (19)
Gaming the measure	The system optimises a score instead of the goal.	Hard rules beside the score (30)
Approval misuse	A person approves one action and another is executed.	Approval bound to the exact action (25)
Silent failure	Nobody can reconstruct what happened or why.	Tracing and audit trails (23)

One rule connects them all, and it is the closest thing this book has to a law:

The control law

As autonomy increases, control requirements increase. The objective is not maximum autonomy. It is useful autonomy under appropriate control.

7. Accountability and governance

An agent cannot be held responsible for anything. A person must be. Research published in Harvard Business Review in May 2026 found, in a large-scale experiment, that describing AI as an employee rather than as a tool shifted accountability away from the people involved^[4]. Treat agents as governed tools with human owners, not as colleagues.

Governance for agents needs a few concrete elements:

- **A named owner for every agent**, who answers for its outcomes and can stop it.
- **A register of agents**: what each one is for, what it may change, which data it may see, and who owns it.
- **Permissions granted per task**, not per agent, and enforced by software rather than by instructions to the model (Chapters 25 and 26).
- **Approvals bound to the exact action**, so that what a person approved is what runs (Chapter 25).
- **An audit trail** that records which decision caused which action (Chapter 23).

- **Governance in proportion to autonomy.** A drafting assistant and an agent that moves money need different controls. Gartner warned in May 2026 that applying the same governance to every agent, regardless of its autonomy and scope, leads to failure^[5].

Regulation is moving in the same direction. Under Article 26 of the EU AI Act, organisations that deploy high-risk AI systems must use them according to the provider's instructions, assign human oversight to people with the necessary competence, training, and *authority*, monitor the system and report risks, and keep its automatically generated logs for at least six months^[6]. As of October 2026 these duties apply from 2 December 2027 for most high-risk uses; check the current dates and whether your use case is classed as high-risk. Note the word *authority*: an overseer who can watch but not intervene does not meet the requirement.

Trust is still scarce. In a survey of 603 business and technology leaders conducted in July 2025, Harvard Business Review Analytic Services found that only 6% of companies fully trust AI agents to run core business processes, and only 12% said their risk and governance controls were fully in place^[7]. Governance is not what slows agents down; its absence is what keeps them out of core processes.

8. Measuring whether it works

A demonstration shows that something can work once. A decision to deploy needs evidence that it works reliably, on cases the system has not seen, at a known cost.

Evaluate at the failure layer

Measure final outcomes, but also the steps that produce them. Otherwise a good score can hide a fragile system, and a bad score can lead to the wrong fix (Chapter 22).

A small set of measures is enough for an executive dashboard:

Measure	What it tells you
Task success rate	How often the agent achieves the goal, on held-out cases.
Cost per completed task	The full cost, including retries, escalation, and review.
Escalation rate	How often it hands work to a person, and whether that is falling.
Intervention rate	How often a person has to correct or stop it.
Incidents	Wrong actions that reached a customer, a system, or a record.
Time to completion	Whether the process is actually faster end to end.

Be wary of measures that can be satisfied without doing the work, and of checks that agree with each other because they share the same blind spot.

Evidence from the book (Chapters 13, 19, and 30)

Letting a model critique its own answer repaired a wrong answer and broke a correct one; only an independent check kept the repair and removed the damage (Experiment 25). A majority vote repaired one error and confirmed another, because the voters shared the same mistake (Experiment 34). An easy-to-measure reward selected exactly the wrong behaviour until a hard rule was added (Experiment 59).

9. People and organisation

Agentic systems change what people do more than how many people are needed. The work moves towards setting goals, defining what good looks like, reviewing exceptions, and improving the system. Three capabilities matter most:

- **Owners** in the business who define outcomes, accept risk, and answer for results.
- **Evaluation** skills: building the test cases and measures that tell you whether an agent works. In the July 2025 survey above, 44% of organisations were prioritising training staff to oversee agents^[7].
- **A platform team** that provides shared models, tools, permissions, logging, and evaluation, so that each new agent does not rebuild the controls.

Prefer open standards for connecting agents to tools and to each other, such as the Model Context Protocol and agent-to-agent protocols (Chapters 10 and 21). They reduce dependence on a single vendor. They do not, on their own, create trust: a standard connection still needs the permissions and checks described in Section 7.

10. Buy, build, or partner

In the MIT study, solutions bought from specialised vendors or built with partners succeeded about 67% of the time, while internally built tools succeeded about a third as often^[3]. Buying is not automatically safer, however. In September 2026 Gartner predicted that by 2028, 70% of enterprises will abandon agentic AI built for them by vendor engineers embedded in their organisation, because costs rise and the customer cannot evolve the system on its own^[8]. Whichever route you choose, make sure your organisation can run, measure, and change the system without depending on someone else.

Ten questions to put to any vendor or internal team:

1. What exactly can the agent change, and what is it forbidden to change?
2. How are permissions granted, and are they enforced by software or by instructions to the model?
3. Which actions require a person's approval, and is the approval tied to the exact action?
4. What is logged, for how long, and can we reconstruct why any action was taken?
5. What evidence shows it works on cases like ours, measured on cases it was not tuned on?
6. What does a completed task cost, including retries and escalations, and how are budgets enforced?
7. What happens when it fails, when a tool is unavailable, or when it is unsure?
8. How does it resist instructions hidden in documents, emails, or web pages?
9. Which open standards does it support, and how would we leave?
10. Who in our organisation will be able to change it after the project ends?

11. A first ninety days

Write the stop criteria before the project starts. A project that cannot fail cannot be evaluated either.

Weeks	Do	Decide
1–2	Choose one or two processes from the “start here” quadrant. Name an owner. Measure today’s cost, time, and error rate.	Is the path variable enough to need an agent, or would a workflow do?
3–6	Build the evaluation first: real, held-out cases and the measures in Section 8. Then build the smallest system that could pass.	Does it beat the baseline on held-out cases, at an acceptable cost?
7–10	Run it in a limited scope, at rung 2 or 3 of the ladder, with budgets, approvals, and an audit trail.	Are incidents, escalations, and interventions within the limits agreed in advance?
11–13	Review the evidence with the owner and the risk function.	Scale, adjust, or stop, against criteria written down in week 1.

12. The executive checklist

Before approving an agent, you should be able to answer every question with a yes.

- ☐ Is the task’s path genuinely variable, so that a fixed workflow would not do?
- ☐ Is there a named owner who will answer for the agent’s outcomes and can stop it?
- ☐ Is the autonomy level chosen deliberately, with the matching controls in place?
- ☐ Can the outcome be checked, and can wrong actions be undone or contained?
- ☐ Is success defined, measured on held-out cases, and compared with today’s process?
- ☐ Is the cost per completed task known, with budgets enforced before actions?
- ☐ Are permissions granted per task and enforced by software?
- ☐ Are consequential actions approved by a person, bound to the exact action?
- ☐ Can we reconstruct what the agent did and why, from its audit trail?
- ☐ Has it been tested against hidden instructions, bad data, and unavailable tools?
- ☐ Do we meet our regulatory duties, including human oversight with real authority?
- ☐ Are the stop criteria written down, and do we have the capability to change the system ourselves?

13. Principle and key takeaways

Briefing principle

Autonomy is a business decision, not a model capability. Grant it only where value can be measured, actions can be controlled, and a person is accountable.

- An agent is a model whose decisions cause actions, with software in control of what is allowed.
- Use the least autonomy that achieves the result; many “agentic” use cases are workflows.

- Start where outcomes are checkable, actions reversible, and volume high; redesign the process rather than attaching an agent to it.
- Manage cost per completed task, and enforce budgets before actions, not after.
- Every agent needs a named human owner, task-scoped permissions, approvals bound to actions, and an audit trail.
- Approve on held-out evidence, not on demonstrations.
- Whether you buy or build, keep the ability to run, measure, and change the system yourselves.

14. Where to go next

The reading path for executives in *How to Use This Book* lists the chapters that matter most for leaders. Each chapter opens with a central question and ends with a principle and key takeaways, which can be read on their own. Every chapter also has a narrated video lecture, and each of the book's 59 experiments can be run and watched step by step in a web browser, with no programming, at <https://agentic-engineering-book.com/lab/>.

Sources

- [1] Gartner. “Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027.” Press release, 25 June 2025. <https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027>
- [2] McKinsey & Company, QuantumBlack. “Seizing the agentic AI advantage: A CEO playbook to solve the gen AI paradox and unlock scalable impact with AI agents.” June 2025. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/seizing-the-agentic-ai-advantage>
- [3] A. Challapally et al., MIT NANDA. *The GenAI Divide: State of AI in Business 2025*. August 2025. As reported in S. Estrada, “MIT report: 95% of generative AI pilots at companies are failing,” *Fortune*, 18 August 2025. <https://fortune.com/2025/08/18/mit-report-95-percent-generative-ai-pilots-at-companies-failing-cfo/>
- [4] M. Kropp, J. Bedard, E. Wiles, M. Hsu, and L. Krayner. “Research: Why You Shouldn’t Treat AI Agents Like Employees.” *Harvard Business Review*, 6 May 2026. <https://hbr.org/2026/05/research-why-you-shouldnt-treat-ai-agents-like-employees>
- [5] Gartner. “Gartner Says Applying Uniform Governance Across AI Agents Will Lead to Enterprise AI Agent Failure.” Press release, 26 May 2026. <https://www.gartner.com/en/newsroom/press-releases/2026-05-26-gartner-says-applying-uniform-governance-across-ai-agents-will-lead-to-enterprise-ai-agent-failure>
- [6] Regulation (EU) 2024/1689 (Artificial Intelligence Act), Article 26: Obligations of deployers of high-risk AI systems. <https://artificialintelligenceact.eu/article/26/>
- [7] Harvard Business Review Analytic Services, sponsored by Workato and Amazon Web Services. Survey of 603 business and technology leaders, July 2025. As reported in N. Lichtenberg, “HBR: Only 6% of companies fully trust AI agents to handle core business processes,” *Fortune*, 9 December 2025. <https://fortune.com/2025/12/09/harvard-business-review-survey-only-6-percent-companies-trust-ai-agents/>

- [8] Gartner. “Gartner Predicts 70% of Enterprises Will Abandon Agentic AI Built by Vendor Forward-Deployed Engineering by 2028.” Press release, 29 September 2026. <https://www.gartner.com/en/newsroom/press-releases/2026-09-29-gartner-predicts-70-percent-of-enterprises-will-abandon-agentic-ai-built-by-vendor-forward-deployed-engineering-by-2028>

Read the complete book

This briefing is part of *Agentic Engineering from First Principles: Building, Controlling, and Scaling LLM Agents in Practice* by [Christiaan M. van der Walt, PhD](#). The complete book follows the journey from a single model call to controlled autonomy in 30 chapters and 59 executable experiments, each printed with its real output.

- **Download the complete book:** <https://agentic-engineering-book.com/#book-download>
- **Watch the video lectures**, one for every chapter: <https://agentic-engineering-book.com/#lectures>
- **Run the experiments in your browser**, no programming needed: <https://agentic-engineering-book.com/lab/>
- **Companion code:** <https://github.com/agentic-engineering-book/code>