GrowPerform and **Serpin**

Getting AI to stick

The hard part of AI is a leadership job, not a technical one, and it decides whether the money you spend ever moves the business. This piece gives you the language to lead it: how to work out what kind of AI you need, four questions for any vendor meeting, and a way to raise the whole thing with your board.

What to take from this

1. Done well, AI gives your best people back the hours they lose to routine work, so they can spend that time on the projects that set the firm apart.

2. The failure rate is high, and the technology is rarely the reason. Getting an agent live is no guarantee it stays live: the AI vendor Sinch reports that 74% of enterprises have pulled a live, customer-facing agent back out of production after a governance failure. The build is the smaller problem. The change and integration work around it is the larger one, and a leader can own it. [1]

3. The firms getting value are the ones building the capability to deploy. They earn trust, then grant the AI more freedom as the evidence comes in, and that lead compounds.

4. Decide how much freedom to give the AI by how reversible the work is, not by how clever the technology can be. The riskier the decision, the less it should act alone.

5. Agree the measure before launch, and give one person inside the business the job of holding it. Understanding comes first, then trust, then use.

6. The real return on a first AI project done well is being able to run the next one faster. That capability, more than any single tool, separates the firms keeping pace from the ones falling back.

A shared vocabulary

Four terms used throughout, so the same words work in a board meeting, a vendor pitch, and the corridor afterwards.

Agentic system

People, AI agents, ordinary software and a redesigned process, working together as one, rather than a tool added on top of the old way of working.

The four freedom levels

Provides information. Recommends actions. Acts with human approval. Acts independently within policy. How much the AI is allowed to do without a person in the loop.

The adoption order

Understanding, then trust, then use. You cannot skip a step, and many firms try to.

The change work

The readiness, trust and redesign around the model that decide whether the investment pays back.

Fuller definitions of these and other terms are in the glossary at the end.

1. The advantage belongs to the curious

Why the money goes in and the business does not move.

Someone talks about agents and orchestration and retrieval, the slide has a diagram with arrows going both ways, and you cannot tell whether what you just heard was clever or empty. So you nod, ask a question that sounds sensible, and keep things moving. Almost everyone in that room is doing the same, and pretending otherwise has become the expensive option.

The intimidation runs in both directions, and that is the part few people say out loud. The business owner cannot challenge the technical team, because the language is foreign. The technical team often cannot see the P&L, the client relationships, or the regulatory exposure the business turns on. Two groups of capable people, each fluent in half the problem, with no shared language between them.

The useful part is this. The advantage now sits with the curious, not the credentialed. You need to know enough to test what you are told: enough to ask the question that shows whether an idea is sound, and to keep asking until the answer makes sense in the terms your business runs on. Your job is to translate between the people who own the goal and the people who build the thing. That is a leadership task, and it is far more learnable than the jargon makes it look.

The stakes are high. Adoption is now near universal and the value is not. In McKinsey's 2025 research, 88% of organisations use AI in some form, yet only 39% report any impact on profit, and most of those see only a marginal effect. [2] A separate BCG study puts the share of companies capturing the large majority of the value at around one in twenty. [3] Spending that hard while so little lands is a fast way to spend money you cannot account for.

Where the value leaks out

Nearly every firm now uses AI.
Only about one in twenty captures its value at scale.

Both figures are shares of all firms. The 5% who capture value at scale sit within the 88% who use AI, so the distance between the two is the value most firms are leaving unclaimed. Sources: McKinsey State of AI 2025 [2]; BCG 2025 [3].

The reasons are rarely technical. The project gets run as a technology rollout instead of a change in how people work, so the tool arrives and the working day around it does not. Underneath sits the real fault line. Nobody owns the join between the business goal and the build, so the builders optimise for something adjacent, and the tool solves the wrong problem well. From the outside, a failed handover and a failed AI look identical, and it is nearly always the handover.

You know the meeting that goes well and then goes nowhere. Everyone leaves nodding, and three weeks later nothing has happened, because the person who needed to carry it forward had no way to sell it to anyone who was not in the room. Interest was never the problem. What was missing was the shared language to carry it past the door. Guidance on AI strategy and on building the thing is everywhere. Guidance on how a leader leads this and makes it hold is not, and that is where the return is decided.

The prize is real and mostly unclaimed. Before the how, one more reason to move.

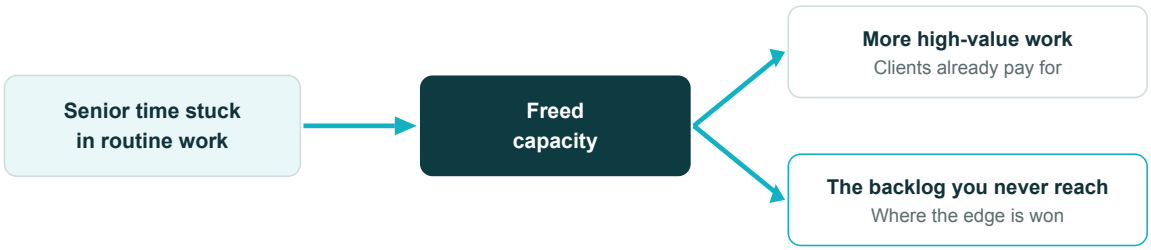
2. What's possible, and the cost of standing still

The prize is capacity, and a group of firms is already taking it.

Start with what a leader actually wants from this, because the prize is larger than a cost cut. The real opportunity is capacity. In most professional and financial services firms, the best people lose a surprising share of their week to work that does not need their judgement: reconciliations, compliance checks, the same figures pulled into the same report every month. That work caps what the firm can take on.

Free that time up, and the question becomes what you point it at. The easy answer is more of the high-value work clients already pay for. The larger answer is the backlog every firm carries and never reaches: the client experience you have meant to fix for two years, or the new service you cannot staff because the people who would build it are the ones stuck in admin. That is where a firm gets its edge, and it is growth you cannot hire your way to.

What freed capacity is for



The firms that free their people and aim that time at the backlog are the ones that reach the edge first.

A firm Serpin is working with in the built environment shows what pointing the technology well looks like. Owner-led and preparing for sale, they knew exactly what drove their value: bids out of the door at a known win ratio. They aimed the technology at that one driver, targeting three times the bid throughput on the same headcount, so the business would be worth more at sale and could run without the founder in every decision.

One driver, chosen deliberately. Bids out of the door at a known win ratio. The target was three times the throughput on the same headcount. The point was a business worth more at sale that could run without the founder in every decision. The discipline is to aim technology at what drives the business, rather than buy something clever and hope.

The firms already moving are pulling ahead

The value on offer is not sitting still, because a group of firms is already capturing it. The lead comes from building the capability, and that capability compounds in a way a one-off purchase does not. Firms deploying now gain real experience at scoping and building. They start with small, contained deployments, earn confidence, then link them across larger parts of the workflow. They grant the AI more freedom as they earn trust in it and improve the output. Each redesigned process leaves cleaner data and reusable controls that the next one inherits, so the early movers move faster with every wave.

The evidence backs the pattern. McKinsey found that redesigning the workflow around AI has one of the strongest links to profit of any change a firm can make, and that high performers are close to three times as likely to have redesigned a workflow. [2] The platforms are arriving too, as the major providers ship governed agents into regulated banking and professional workflows. [4] The advantage is available now, and it grows for the firms that act on it.

Three lenses for any AI move

Possible. What the technology can now do reliably, which is whole pieces of work from start to finish.

Desirable. Whether it serves a value driver you already own: capacity, growth, resilience, reach.

Practicable. Whether the people, process and data are ready to carry it.

The prize is real, and some firms are already ahead. The question a leader is left with is how to lead it.

3. How a leader leads this

One idea threads the whole toolkit.

You are not the one building the AI. Your job is to lead the effort around it, and that comes down to four decisions, taken roughly in order over the life of an AI move. The four tools that follow are one method rather than four separate topics, and this is the idea that connects them. How much freedom the AI has to act, its freedom level, is the setting that everything else follows from. Choose it, and readiness, the change work, and the questions you ask all follow from that choice.

You choose the level a given job actually needs, and that choice sets the bar. A claims agent that acts with human approval asks far more of your people, your data and your controls than a research agent that only informs, so the same firm will start one in weeks and the other in months. Across jobs and over time, you climb: a firm earns trust on the lower levels, then takes on higher-freedom work with that experience behind it. The setting is chosen per job. The climb happens across the programme.

The setting, and what follows from it

Choose the freedom level first, because it decides how much the next two ask of you. Readiness and the change work both scale with it. The four questions turn all three into things you can say out loud in a room.

This idea links back to the firms pulling ahead, who climb these levels as they earn trust. Your part throughout is to keep translating between the business goal and the technical work until the answer makes sense, whether the firm is weighing a single tool or a dozen.

4. How much freedom to give the AI

The question is how much it may do without a person in the loop.

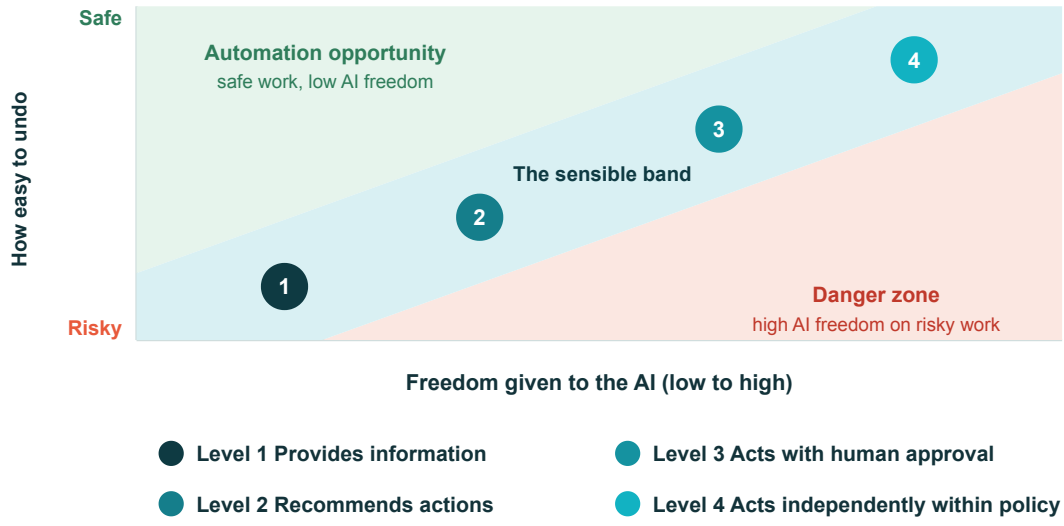
Different things get lumped under one word here, so start with the question that sorts them. The question is how much freedom you give the AI to act without a person in the loop. The level you choose is the setting the rest of the toolkit follows from. Serpin sorts the work into four levels. At the first, the AI does not act at all. It reads and flags, and a person does everything from there. At the second, it drafts what it would do and waits for approval. At the third, it finishes the work on its own, with people reviewing afterwards. At the fourth, it runs inside limits you set, with no one signing off each action.

Level	What the AI does	What people do	Example	What it asks of your firm
1 **Provides information**	Produces information for a person to use	Act on the information	A research briefing agent	A named owner and spot-checks. Ships in weeks.
2 **Recommends actions**	Proposes specific actions for a person to take	Take the action	A contract review agent	An owner and an AI lead, monthly review. Six to ten weeks.
3 **Acts with human approval**	Performs actions, gated by a person's sign-off	Approve before anything happens	A claims agent	An owner, an AI lead and an approver pool, quarterly review. Three to six months.
4 **Acts independently within policy**	Acts on its own inside an enforced policy	Review by exception	A fraud screening agent	Executive or board ownership, continuous monitoring, a named stop protocol. Six to twelve months.

Levels two and above are where the discipline bites. The approval level is for anything you cannot easily undo, or that a regulator will ask you to account for later. If your team learns to wave the drafts through unread, you have the look of a control and none of the substance. The hands-off level is only for safe, high-volume work that is easy to verify and cheap to get wrong. Almost nothing that touches a client's money or a regulated decision belongs there, and a vendor who wants to put that work there should say exactly why.

This next point runs against the instinct a demo is built to trigger, and most vendors will not lead with it. The right question is how little freedom the job safely needs.

The riskier the work, the less freedom you hand over



Match the AI's freedom to the work. Safe work that is easy to undo can run at a higher level. Risky, hard-to-undo work stays at the lower levels, with a person in the loop. Often the honest answer is that you do not need AI at all, because a well-built spreadsheet is cheaper and easier to trust than a model.

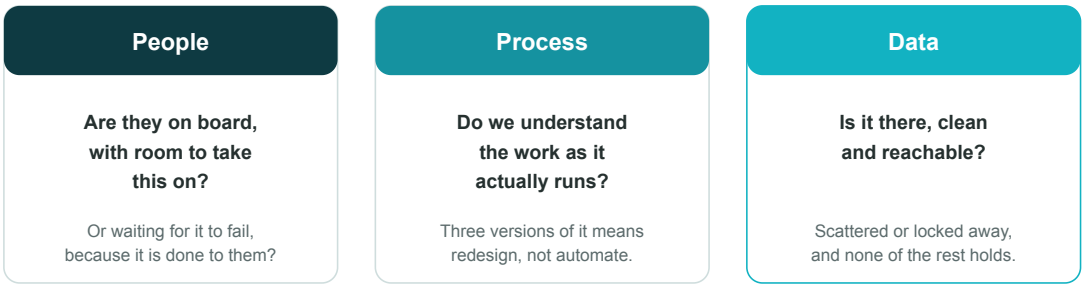
The freedom level you choose decides how much the rest of the work demands. A higher level of freedom needs more readiness, more trust and tighter control. So the next question is whether your organisation is ready for the level of freedom your use case warrants.

5. Are you ready? People, process, data

Readiness is read against the level you chose, and it is where the change work starts.

Readiness is not a box to tick, and it is judged against the level you just chose. The more freedom the AI will have, the more your people, process and data have to be in shape before you build. Starting well rests on two honest answers. The first is a specific business goal, stated as a value driver you already care about: win more bids, cut the time to onboard a client, catch a category of risk earlier. "We want to use AI" is not a goal. A good answer here is concrete and slightly uncomfortable, because it commits you to something you can measure. The second is an honest read on whether you are ready.

The three honest questions



Take the people first, because they carry the change or quietly sink it. Then the process, which is where readiness most often falls down. If three people describe the same process three ways, you are ready to redesign it, not to automate it, and there is no point making a mess run faster. Then the data, because if it is scattered, dirty or locked away, none of the rest holds.

Working through these three questions is the first half of the change work, and it is the job of a discovery stage.

How the discovery stage answers the key questions

The discovery stage is the first stage of Serpin's Agent Discovery and Design method. The discovery stage answers the people, process and data questions with evidence, before any budget is committed, and it runs in a set order.

1. **Name the problem, and what a good outcome is worth.** Who is affected, what the problem costs the firm today in time, money and risk, and the value of fixing it. For the built-environment firm, the problem was the bid process: a known cost per bid, a known win ratio, and a target of three times the throughput on the same headcount.

2. **Test the idea against the alternatives.** Often a simpler route, better-organised code or a well-built spreadsheet, serves better than an agent. Saying so is part of the discipline.

3. **Build the business case at the real running cost.** An AI system charges every time it runs, and that charge rises with volume, so a case built on the build cost alone fails its first serious review. The discovery stage models the cost at real volume and checks that the data the system needs is reachable and usable.

4. **Define what good output looks like.** Capture examples of excellent output, judged by the people who do the work today, so the build has a clear standard to aim at and to be tested against.

5. **Map the process as it really runs.** Including the unwritten rules that live only in people's heads. This is where most readiness gaps hide, and it is why redesign usually beats automation.

6. **Decide: go, or no-go.** The output is a decision a finance director can defend: a validated case, an honest read on readiness, and a plan for the change work, or a clear reason not to proceed.

Run this way, the discovery stage is the fastest and cheapest way to find out whether a project is worth starting.

A good discovery stage gets you to a build worth doing. A build worth doing still fails if it does not hold, and holding it is the harder half.

6. Making it stick

The half that looks least technical, and decides the return.

Making it stick is what turns a build into a return, and it gets the least attention because it looks the least technical. It is also where the freedom level matters one last time. The higher the freedom level, the more trust you have to earn and the more of the change work you have to do. The change work runs across three layers, and it is what turns a working model into a working agentic system: people, software and a redesigned process holding together in the real job.

1

Absorb the change

The organisation can take the change on without breaking. That means the capacity to do it, and a plan for the ordinary curve of any new way of working: anxiety, then curiosity, then frustration, then a settled competence. Scale the effort to the risk, and the four levels tell you how much.

2

Adapt behaviour

The new way of working actually holds. That means the measure set before launch, one owner inside the business, and the adoption order respected: understanding, then trust, then use.

3

Build a habit of change

The firm gets better at the next change, not only this one. Done well, the change work speeds a deployment and helps it hold, and it leaves behind a repeatable way to run the next AI decision.

Keep what is strong and distinctive in your culture. Adapt or introduce change only where you actually want it.

Set the measure before you launch

Decide, in plain business terms, what this AI is meant to move, and agree it out loud with the team before anyone flips the switch. Name the number the business runs on: bids won, or errors caught before they reach a client. "Usage went up" and "the model performs well" are not that number. Set the measure after launch and you will

measure whatever the tool happens to report, which is how firms come to feel informed and stay blind. Few set it first. In professional services, only around 18% track the return in any form. [5]

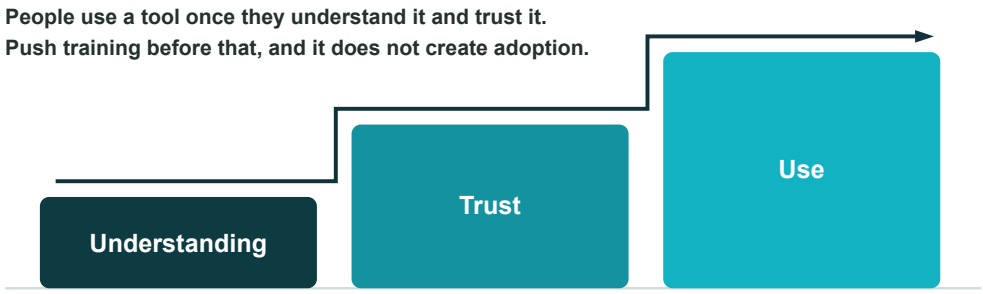
Name an owner inside the business

Name one from day one. Not the vendor, not the consultant, and not a committee, but one person whose job is to hold the join between the goal and the tool once the builders have gone. The translation gap does not close when the technology is delivered. Somebody has to keep asking whether the tool is still solving the right problem. When that person exists, the change has a home. When they do not, it drifts back to how things were, and nobody notices until the return has gone.

Understanding, then trust, then use

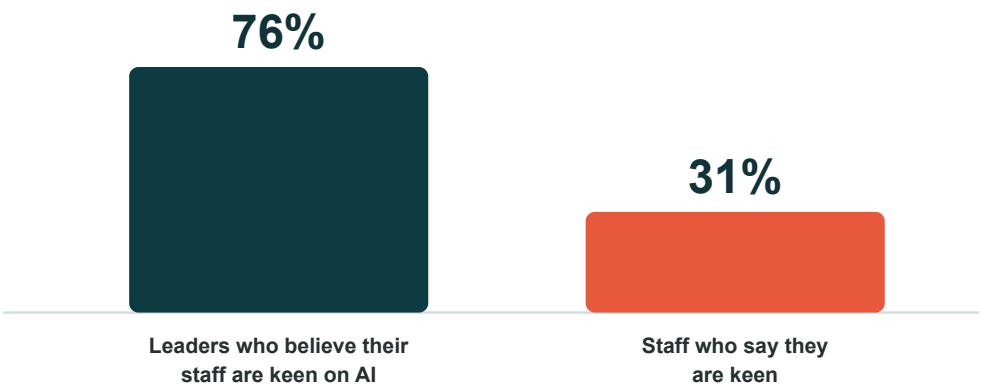
The order matters, and the instinct runs the other way. The reflex is to train people on the tool and expect them to adopt it. It does not work like that. People have to understand what the tool is for and why, in words that make sense to them, before they will trust it, and they will not properly use something they do not trust, whatever the mandate says.

Understanding comes first, then trust, then use



Whether people are truly on board matters more than leaders tend to think. In a 2025 survey, 76% of executives believed their staff were enthusiastic about adopting AI, while only 31% of the staff were. [6] Leading from an assumption that far off means running a project you cannot see.

What leaders think, and what staff feel



Most leaders are managing to a picture more than twice too rosy.

Source: HBR 2025 [6].

One more thing separates a change that sticks from one that fades. Redesign the job around what the tool now does, rather than retrain people to do the old job with a new tool. The point of buying the capacity back was to put it into work that matters, and that only happens if someone deliberately points it there.

These three layers are the method. The last tool makes them portable.

7. The four questions to walk in with

Print this. Take it into any vendor or internal meeting.

Here is the practical part to keep. The decisions above, turned into four questions in a leader's plain voice, each with a sense of what a good answer sounds like against a worrying one.

1 What could this make possible that we already care about?

Good	Names your value driver back to you, such as more bids won or a category of risk caught earlier.
Worrying	Stays on the technology and never quite arrives at what changes for the business.

2 Are we honestly ready, and where are we not?

Good	Candid about where the gaps sit, in the people or the data, and treats them as work to do.
Worrying	Presents everything as ready and easy. The gaps you are not told about surface later, at the worst time.

3 How will we know it is working, in terms the business understands?

Good	Offers a measure you already use, agreed before launch.
Worrying	Offers usage figures and model scores, which show the tool is touched, not that it is paying back.

4 Who owns this once the builders are gone?

Good	Names a real person inside your business, and describes how the knowledge transfers to them.
Worrying	Assumes, without quite saying so, that the vendor or consultant stays forever.

How to raise it with your board or your CEO

There is an AI move here tied directly to **[our value driver]**. I want us to be honest about whether we are ready before we commit, both our people and the data underneath, and to agree upfront how we will measure whether it is paying back, in business terms.

And I want a named owner inside the business who keeps hold of it after the build is done. If we can answer those, it is worth doing. If we cannot yet, I would rather know now than after we have signed.

A script you can lift, in the terms your board already uses.

8. What changes for the leader

You come out with something more valuable than the tool.

The leader who walked into that first meeting lost can walk into the next one with the language, the questions, and a sense of what good looks like. That is the difference between leading the room and nodding along in it. When the shared language is there, the team understands what the tool is for, so they trust it, so they use it, and the capacity you freed goes into work that grows the business.

You also come out with something more valuable than the tool itself: a repeatable way to run the next AI decision, rather than starting each one over. That matters, because the technology will keep changing, and faster than most leaders expect. The firms that pull ahead will not be the ones that picked the cleverest tool this year. They will be the ones that built the capability to take on the next change, and the one after, without a fresh scramble each time. Doing the first project properly is how that capability is built. The job underneath it stays the same. Point the technology at something that matters, and make the new way hold. That work is yours, and you can learn it.

What to do now, and it needs no budget

The capability that decides whether an AI investment pays back is the change work around the model. The next step is small. Take the four questions to the people who would build your next AI move, and decide honestly whether you are ready.

If the honest answer is that the readiness or the change work is where you are least sure, that is the work we do. A short, structured discovery is the fastest way to find out whether a project is worth starting, and it is where a conversation with us begins.

This is the first of a short series on leading AI change. A companion piece looks at governance, the controls and audit trail that make a higher-freedom system defensible to a regulator.

About the authors

Rebecca Waker

GrowPerform

Rebecca leads GrowPerform, a change-management practice for AI projects. She spent around a decade in banking and fintech, in operations and organisational change, including corporate change and transition work at Barclays. GrowPerform handles the people, process and adoption side of an AI project: the readiness work before deployment, and the launch and adoption work afterwards, so the investment in the AI pays back.

Scott Druck

Serpin

Scott is a co-founder of Serpin, an AI consultancy for mid-sized professional services and regulated firms. He has spent his career running change and transformation programmes inside regulated organisations, and now designs, builds and governs AI agents for financial and professional services. Bounded Agency, his framework for governing AI agents in regulated industries, underpins Serpin's work on higher-autonomy systems.

9. Glossary

Plain definitions of the terms used in this guide.

Generative AI

AI that creates new content, such as text, images or speech, rather than only finding patterns in existing data.

Agent

A language model put to work on a task, able to use the firm's systems and data, produce finished work and make decisions, rather than only answer a question.

Tools

The systems and data an agent is allowed to use, such as a database, a CRM or an email system.

Fabrication

When a model produces something that reads convincingly but is untrue, such as an invented figure or a made-up source. Also called hallucination.

Running locally

Keeping the AI and its data inside the firm's own environment, rather than sending them to an outside service. Also called on-premise.

Language model

A system that works with text: reading, writing, summarising and reasoning. ChatGPT and Claude are examples.

Agentic system

A production system of one or more agents wired together with code, controlled access to data and systems, and the checks that keep output reliable.

Freedom level (autonomy level)

How much a system does without a person, from providing information to acting on its own within an enforced policy.

Grounding

Tying a model's output to real, cited sources so each claim can be checked. The main defence against fabrication.

Responsible AI

Building and running AI to defined standards of accuracy, fairness, transparency and accountability.

References

1. Sinch, *The AI Production Paradox*, May 2026 (survey of 2,527 senior decision-makers across 10 countries, including financial services). 74% of enterprises have rolled back or shut down a live AI customer-communications agent after deployment following a governance failure; 62% have agents in production; 98% are increasing AI investment in 2026. sinch.com (https://sinch.com/news/sinch-releases-ai-production-paradox/)
2. McKinsey, *The State of AI in 2025*, November 2025. 88% of organisations use AI in some form; only 39% report any EBIT impact, most attributing under 5% of EBIT to it; workflow redesign is among the strongest predictors of impact, and high performers are close to three times as likely to have redesigned a workflow. mckinsey.com (https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai)
3. BCG, *Are You Generating Value from AI? The Widening Gap*, 2025. Around 5% of companies capture the large majority of the value on offer from AI. bcg.com (https://www.bcg.com/publications/2025/are-you-generating-value-from-ai-the-widening-gap)
4. Deployment maturity, 2026: Fiserv agentOS and Experian's Agent Operating System bring governed agents into banking; Anthropic, OpenAI and Google are shipping ready-made agents for underwriting, compliance and related regulated work. Sources: Fiserv investor release (14 May 2026); Experian newsroom (2 June 2026); Fortune coverage of Anthropic finance agents (5 May 2026).
5. Thomson Reuters Institute, *2026 AI in Professional Services Report*, February 2026. 18% of organisations track the return on AI tools in any form. thomsonreuters.com (https://www.thomsonreuters.com/en-us/posts/technology/ai-in-professional-services-report-2026/)
6. Harvard Business Review, *Leaders Assume Employees Are Excited About AI. They're Wrong*, 2025 (BCG and Columbia Business School survey of 1,400 US employees). 76% of executives believe employees are enthusiastic; 31% of individual contributors say they are. hbr.org (https://hbr.org/2025/11/leaders-assume-employees-are-excited-about-ai-theyre-wrong)