bosio digital

GUIDE
2026 EDITION

Where to put your *first* AI agent

Most first agents fail on selection, not on technology.

TIME NEEDED
20 minutes and your process list

YOU LEAVE WITH
One process, scored and chosen

FROM
bosio.digital AI consulting

BEFORE YOU START

The technology is rarely the reason it did not work.

A capable system pointed at the wrong process produces an expensive demonstration and no return.

MIT's NANDA researchers looked at an estimated $30 to $40 billion of enterprise generative AI spending and found roughly 5% of integrated pilots extracting measurable value. Their diagnosis did not blame the models. The systems, they wrote, "do not retain feedback, adapt to context, or improve over time." Gartner expects more than 40% of agentic AI projects to be cancelled by 2027.

Those numbers get quoted as a warning about AI. They are better read as a warning about selection. The same system pointed at the right process pays for itself before anyone has to defend it in a budget meeting.

This guide is about that choice, and only that choice.

Four steps, and you can finish them this week

Step 1

Settle what an agent is, so you are not sold one where a simpler thing would work better.

Step 2

Score your own processes against four traits. Page 6 is the scorecard.

Step 3

Compare your top scorer against five starting points that are known to pass.

Step 4

Put one name on it, and set the measure before anything gets built.

WHAT YOU WILL FIND

The process that scores highest is almost never the one that was most discussed in the leadership meeting. It is usually smaller, more boring, and touches more people than anyone expected.

A chatbot is a tool you use. An agent is *work you delegate.*

Three properties have to be true at once. Most of what is sold as agentic fails at least one of them.

1 It takes multiple steps

It moves toward a goal without being re-prompted at every stage.

2 It uses tools

It can search, call a system, write a record, run code. Not only produce text.

3 It adapts

If step three fails, it notices and tries something else.

WHAT IT IS	WHAT IT DOES	WHERE IT STOPS
A chatbot	Answers turn by turn	No goal of its own, takes no action in your systems
Generative AI	Produces content when asked	Reactive. Generates, then waits
RPA	Follows a fixed script in your systems	No judgment. Breaks when the screen changes
An agent	Decides, acts, checks, continues	Needs delegation, and therefore management

WHY THE DISTINCTION MATTERS TO YOU

You already know how to manage delegation, because you do it with people. You do not hand a new hire the company card and full system access on day one. You scope what they can do, you check the work until you trust it, and you keep the accountability for the outcome. Everything difficult about agentic AI follows from that one word.

Does the task actually need an agent?

Anthropic's engineering team draws a line worth adopting word for word.

A **workflow** is a system where models and tools are "orchestrated through predefined code paths." An **agent** is a system where models "dynamically direct their own processes and tool usage." A workflow runs on rails you laid down in advance. An agent decides the route.

Their own advice is to start with the simplest thing that works, a single model call or a fixed workflow, and to escalate to real autonomy only when the task demands it.

That advice cuts against most of what you will be shown. Many of the highest-value wins available to a company of 20 to 200 people are fixed workflows, not autonomous agents. They are cheaper, more reliable, and far easier to govern.

THE QUESTION TO ASK A VENDOR

Does this task actually need an agent, or is this autonomy I will have to supervise?

If the honest answer is the second, you have just saved yourself the most expensive part of the project, and it is the part nobody budgets: the supervision.

A NOTE ON SEQUENCING

Choosing a workflow over an agent is not a smaller ambition. It is the same ambition with a shorter path to evidence. The companies that end up running real autonomy almost all got there by shipping something simpler first.

Four traits, scored yes or no.

They are deliberately binary. A process that scores four is a first project. A process that scores two is a second-year project wearing a disguise.

1 It touches more than one person

Work that lives entirely inside one person's day produces a private time saving and no organizational evidence. Work that crosses a handoff produces something other people can see, which is what makes the second project easier to fund than the first.

2 The right answer follows rules the team can state out loud

Not rules that are written down. Rules someone can say in a sentence if you ask. If two experienced people would disagree about the correct output, the process is not ready. That disagreement is judgment, and judgment is the thing you are keeping.

3 Nobody has to learn a new behavior

Every new habit you ask for is a place the project can quietly stop. The best first candidates change what happens to the work, not what the person does.

4 It runs inside tools the team already opens

If the output arrives somewhere nobody looks, it did not arrive. A result that appears where the work already happens gets used without anyone being asked to use it.

AND A FIFTH TEST, WHICH IS A VETO

A mistake has to be cheap to catch. If a wrong answer can reach a customer, a regulator or a bank account before a human sees it, this is not a first project regardless of what it scored on the other four.

Your processes, against the screen.

Write down every process a leader has complained about in the last quarter. Then score each one. One point per yes.

YOUR PROCESS	CROSSES MORE THAN ONE PERSON	RULES CAN BE SAID OUT LOUD	NO NEW BEHAVIOR	TOOLS ALREADY OPEN	MISTAKES CHEAP TO CATCH	TOTAL

Reading the result

5 of 5	Start here. Do not keep looking for something more impressive.
4 of 5	A good first project, if the veto passed. Name the missing trait and decide who closes it.
3 or below	Not first. Write it on the second list and leave it there.
Veto failed	Not a first project at any score.

IF NOTHING SCORES ABOVE THREE

That is a finding, not a failure. It usually means the work that matters here runs on judgment rather than rules, which is worth knowing before you spend anything. The place to look next is the handoffs between functions, rather than the work inside them.

Five starting points, and they are all *unglamorous.*

Each is scored in our published use case guide on implementation complexity out of five, and on time to value. They pass the screen for most companies, which is exactly why they look ordinary.

STARTING POINT	COMPLEXITY	TIME TO VALUE	WHY IT PASSES
Meeting summaries and action tracking	**1 of 5**	**1 to 2 weeks**	Crosses everyone, rules are statable, runs inside the meeting platform already in use
Knowledge base upkeep	**2 of 5**	**2 to 4 weeks**	The rules are the style guide you already have, and nobody changes what they do
Conversation summarization and analysis	**2 of 5**	**2 to 4 weeks**	Reads what already exists, and mistakes are visible immediately
Ticket routing and prioritization	**3 of 5**	**4 to 6 weeks**	Rules can be stated, crosses the whole support function, wrong routing is cheap to fix
Document processing and extraction	**3 of 5**	**6 to 8 weeks**	High volume, checkable output, sits in the system the work already lives in

WHAT IS MISSING FROM THIS LIST, ON PURPOSE

Candidate screening and contract risk analysis both score four of five on complexity and take eight to ten weeks, and both fail the veto for most companies, because a wrong answer reaches a person or a counterparty. They are good third projects. They are bad first ones.

WHAT THE EASIEST OF THESE RETURNS

Meeting documentation typically removes 80 to 90% of manual note taking. That is not a transformation. It is a visible, checkable result inside a month, which is the only thing that makes the second project possible.

One process.
One owner.
One measure.

One process, not a portfolio

The failure pattern has a name, agent sprawl, and its signature is that the fifth agent makes the company less intelligent than the first one did, because nothing they learn is shared and nobody can say what any of them is for.

One owner, a named person

Not a committee, and not a department. The next page is about who.

One measure, recorded before you start

The most common reason a working project cannot be defended is that nobody wrote down what the before looked like. Record the current state this week: how long it takes, how often it is wrong, how many people touch it. Fifteen minutes of measurement now is the difference between a result and an anecdote.

WHAT THIS LOOKS LIKE WHEN IT GOES THE OTHER WAY

VISUAL FACILITATORS, a Hamburg firm working worldwide in visual facilitation and graphic recording, had adopted AI early and often. By the time we started, 40 tools, skills and agents were running side by side, with no shared context and no quality floor. The work was not short of AI. It was short of a system.

90 → 30	**+50%**	**40 → 1**
minutes to a first proposal draft, and sharper	content output, 15 pieces a week, no new headcount	scattered tools and agents into one owned system

Every output passes a quality check before it ships. The firm owns and extends the system itself, on its own European storage, with no dependency on us. The order is the lesson: the result did not come from the fortieth tool. It came from deciding what the one system was for.

WHAT TO MEASURE AFTERWARDS

Not seats activated, and not logins. A real task, done the new way, by someone who was not asked to do it that way. That is the only adoption number that has ever predicted anything.

THE ORDER THAT DECIDES THE OUTCOME

Most companies start by buying agents.

The sequence that works starts by building the conditions agents need.

1 A map

One real process, understood well enough to redesign. Dropping an agent onto a broken process automates the breakage faster. This is also where the workflow-or-agent decision actually gets made.

2 Context and structure

An agent is only as good as what it knows about your business and how clearly its job is defined. The companies in MIT's successful 5% had built the memory and context that let their systems improve. The 95% had not. This is the difference between a system that starts every task as a stranger and one that compounds.

3 Governance before autonomy

Start every agent at approver level: it proposes, a human releases. Watch it. Grant more independence only as it earns it. Cap what it can spend and what it can touch. Keep a human above the loop, not in it.

6%

QUALIFY AS AI HIGH PERFORMERS, WITH MORE THAN 5% EBIT IMPACT

40%+

OF AGENTIC AI PROJECTS EXPECTED TO BE CANCELLED BY 2027

McKinsey's 2025 survey found 88% of organizations using AI in at least one function, about a third beginning to scale it, and 6% qualifying as high performers. Do the three moves in that order and you are building what those companies built. Reverse it, buy first and scramble for context and governance after the invoice arrives, and you are building what Gartner expects to be cancelled.

The technology is identical in both stories. The sequence is the whole difference.

THE OWNER

Whoever owns the work owns the adoption of AI into it.

The reflex in most companies is to hand this to whoever signed the software agreement. That reflex is the most common structural cause of a stalled rollout.

One accountable name at the top. In most mid-market companies that is the CEO. Committees advise. One person answers, the way the CFO answers for the numbers. Then four lanes, none of them optional.

LANE	OWNER	WHAT THEY OWN
The why	CEO	The story and the standard, said plainly and demonstrated in the open
The how	Function leaders	Adoption inside their own workflows. Non-delegable, because the friction is only visible from inside the function
The rails	IT	Identity, access, data boundaries, model strategy, security, governance
The daily loop	A named enabler	Protected time, a monthly deliverable, a mandate co-signed by their leader

THIS IS NOT A DEMOTION FOR IT

As autonomy arrives, the rails become the most consequential work in the building. Deloitte finds about 75% of enterprises planning agentic deployments within two years, and only 21% reporting a mature governance model for them. The measure for IT changes from "nothing bad happened" to "how fast can a function safely ship on our rails."

THE ROLE MOST COMPANIES SKIP

The enabler is what decides whether usage compounds or plateaus at the enthusiasts. Four hours a week and one deliverable a month is enough. A fan with a hobby is not.

FIVE FAILURES, IN THE ORDER THEY ARRIVE

Where this goes wrong.

Technological overreach

Advanced capability bought before the basics exist. The screen on page five is the cheapest protection against this.

The human element treated as a phase two

Training and process redesign left until after the build. By then the habits have set around the old way.

No baseline

This is the one that costs you the second project rather than the first. A working result with nothing to compare it to is an anecdote.

Fragmented starts

Several disconnected projects at once, which produces integration problems and no shared learning. One, then the next.

Delegation confused with abdication

The most instructive case in the category is a fintech whose AI customer service assistant handled 2.3 million conversations in its first month, took roughly two thirds of chat volume, and cut average resolution from eleven minutes to under two, with about $40 million in projected profit improvement. Then the company acknowledged it had cut human staff too aggressively, and began rehiring specialists for the complex, ambiguous and emotionally charged cases.

THE LINE THIS GUIDE IS REALLY ABOUT

That is not a story about AI failing. It is a company finding out where the line between delegation and abdication actually sits, and having to walk back across it. Knowing where that line runs through your own work, before you build, is most of what this guide is for.

WHAT TO DO NEXT

Score one list. Name one person.

If you own the business

Run the screen on page six yourself, with your own list, before you ask anyone else to. Then put one name against the winner, and ask that person what the current state measures. If they cannot answer, that is the first week of work.

If you run operations

The handoffs between your function and the next one are where the highest scorers usually hide. Score three of them this week.

If you are the person who will run it

Ask for the two things that make the role real: protected time, and one deliverable a month that someone reads. Without both, this becomes a hobby.

Where this came from

Everything in this guide is published in full, and free, at bosio.digital: the fifteen AI use cases scored for mid-market companies, agentic AI explained for the mid-market leader, who owns your AI adoption, and AI agent sprawl.

Other guides in this set

The AI Budget Worksheet, on what an engagement actually costs in year one. The context engineering how-to, on the five things your AI cannot work out on its own. Twelve questions to ask an AI consulting firm before you sign.

IF YOU WANT A SECOND PAIR OF EYES ON YOUR SHORTLIST

Send it over, or take 30 minutes with us. If we are not the right fit for where you are, we will say so, and point you somewhere better.

bosio.digital/contact

Sources: MIT NANDA on enterprise generative AI pilots. McKinsey, The State of AI, 2025. Deloitte enterprise agentic AI research. Gartner forecast on agentic AI project cancellations. Anthropic engineering, on the distinction between workflows and agents. IBM, on the perceive, reason, act, learn loop. Klarna public reporting and subsequent statements. Client figures as published at bosio.digital/clients.