SERPIN

AGENT IMPLEMENTATION GUIDES

Serpin First AI Agent Playbook

A simple guide to picking your first AI agent project. For product, transformation, and operations leaders.

CHOOSE YOUR FIRST PROJECT

What is in this guide.

AI agents are powerful. Picking the right first project matters.

An AI agent can take on work that consumes significant skilled time across your firm. Drafting documents from known inputs. Pulling information together across systems. Reviewing material against defined rules. The kind of repetitive but skilled work that eats hours of senior people's days.

AI agents are different from traditional software. They make judgements. They take actions. They use data drawn from across your business. That brings genuine complexity, and means an agent needs thoughtful design, deliberate controls, and clear ownership to run reliably.

> Not every potential project is the right one to start with. The first AI agent your firm builds matters more than the second or the third, because it sets the trajectory for everything that follows.

A well-chosen first project ships, delivers value, and builds your firm's confidence with AI agents. A poorly-chosen one stalls, consumes time and budget without producing a usable result, and makes the next project harder to get approved.

This guide helps you choose well. It walks you through a small set of questions about a project you have in mind, surfaces what it would actually involve, and gives a clear view of whether it is a strong first choice.

The guide is built on **Bounded Agency**, a way of designing AI agents that Serpin developed for reliability, security, and value. It draws on research into what makes production agents work, combined with our own experience building production agents.

Two stages of thinking, one clear outcome.

Two stages of work that lead to a clear answer about the project you have in mind.

WHAT YOU WILL DO

STAGE 1

Business and use case

Six questions that surface what the agent would do, the value it would deliver, the systems it would touch, and the consequences of an error.

01 What is the work the agent would do?
02 What areas of risk does the work touch?
03 What kinds of decisions will the agent make?
04 What systems and data will the agent use?
05 Where does the agent's output go?
06 What happens if the agent gets it wrong?

STAGE 2

Level of autonomy

Your answers place the project at one of four levels. The level tells you what controls, oversight, and ownership the agent needs.

- **Provides information.** Drafts content for a person to use.
- **Recommends actions.** Proposes actions for a person to take.
- **Acts with human approval.** Each action signed off before it commits.
- **Acts independently within policy.** Operates inside an enforced policy.

WHAT YOU GET

A clear picture

A clear view of the project's value, what it would commit your firm to, and whether it is your strongest first choice.

- The level of autonomy the project would sit at.
- The ownership, oversight, and controls that level needs.
- Whether this is your strongest first project.

Two practical notes.

What we will not cover at this stage

At this stage we are not going to get into agent design, technical specification, or build work. That comes later, once you have decided which project is the right place to start. This stage is about choosing the right project to begin with. The choice is the most important decision in the whole process, which is why it deserves its own short, structured guide.

If you do not have a project in mind yet

You do not need a fully formed plan. If a possible AI agent project is already on your mind, use that. If you are starting from a blank page, think about a business process inside your firm that takes significant human effort today, particularly one where the work is repetitive but requires skill. That is usually the type of process where an AI agent can add the most value early on.

> **Ready to begin.** The next section helps you recognise what makes a strong first AI agent project, before you work through the six questions.

What makes a strong first AI agent project.

The best first project delivers real value while being achievable for a firm still learning how to build and run agents.

Look for high value, achievable work

Strong first projects sit in **labour-intensive processes** that consume skilled time but follow clear steps. Drafting documents from known inputs. Pulling information across systems. Reviewing material against a set of rules. An AI agent can take this work on quickly, freeing your team for higher-value work.

Treat the first one as the foundation

The aim of the first project is to deliver useful value *and* to build your firm's confidence with AI agents. A shipped first agent unlocks the second. A stalled one makes the second harder to get approved.

What to avoid as a first project

Some projects are poor first choices, even when valuable. As a rule, avoid:

- Complex data needs, or interfaces with multiple systems.
- Actions taken with other parties without a human in the loop.
- Business processes that are not very well understood inside the firm.
- Heavy regulatory or audit obligations not yet worked through with an AI lens.

These can still be great agentic opportunities. They are just rarely the right place to start.

Six questions about your project.

Hold one project in mind as you work through these. For each, the guide explains what it surfaces, how to answer it well, and shows an example.

QUESTION 01

What is the work the agent would do?

Why this matters. A specific description of the work is the foundation for everything that follows. With it in place, decisions about value and complexity become concrete.

How to answer it well. Describe the business process or task the agent would do. Name who does this work today, how often, and what the firm would gain.

> EXAMPLE ANSWER
>
> "Drafting client-meeting briefings from our internal knowledge base and external research. A senior consultant currently spends 4 to 6 hours per briefing across around 200 briefings per quarter. We want to reduce prep time and free senior capacity."

QUESTION 02

What areas of risk does the work touch?

Why this matters. Different kinds of work carry different risks. Knowing which categories your project touches tells you what controls would sit around the agent later.

How to answer it well. Five categories cover most projects: regulatory, financial, reputational, customer-facing, operational. Most touch two or three. Name them, plus any regulatory framework relevant to your sector.

> EXAMPLE ANSWER
>
> "Reputational, because briefing quality affects client meetings. Confidentiality, because the knowledge base contains client-sensitive material. Operational, because briefings sit on the engagement team's critical path."

QUESTION 03

What kinds of decisions will the agent make?

Why this matters. AI agents differ in how much decision-making they take on. Where the decision-making sits determines the level of autonomy your project would have. This is the question that, more than any other, points you toward the right answer in Stage 2.

How to answer it well. Be specific about where decisions sit. Will the agent produce information for a person to act on, recommend specific actions, carry out actions with sign-off per action, or act inside a defined policy? Aiming higher than needed leads to over-engineered controls; aiming lower leads to problems in production.

> EXAMPLE ANSWER
> "The agent produces a draft briefing for the consultant. The consultant decides what to keep, edits it, and signs off the final version before it is shared with the client."

QUESTION 04

What systems and data will the agent use?

Why this matters. The systems an agent reads from, the systems it writes to, and the sensitivity of the data involved all shape how much design work is needed around access and protection.

How to answer it well. List the systems the agent will read from, the systems it will write to or act on, and the sensitivity of each. Customer personal data, payment data, employee data, commercially confidential material, and regulated data each carry specific obligations.

> EXAMPLE ANSWER
> "Read-only access to the internal knowledge base (client-confidential) and public research feeds. No write access; output goes to the consultant as a draft document for review."

QUESTION 05

Where does the agent's output go?

Why this matters. Output that stays inside the team is much lighter to design around than output reaching customers, regulators, or the public directly. Reversibility matters too: if the output is wrong, can the firm correct it, or is the action already committed?

How to answer it well. Identify where the output goes and how reversible it is. Internal use only is lightest; customer-facing is heavier; regulator-facing or public is heaviest. Then say whether the action can be pulled back, or is committed once taken.

> EXAMPLE ANSWER
>
> "The draft briefing is sent to the consultant for review and edit. Nothing leaves the firm until the consultant approves it. Fully reversible at the review stage."

QUESTION 06

What happens if the agent gets it wrong?

Why this matters. An honest view of the consequences of error calibrates the depth of design and oversight the project needs. Some errors are caught quickly by the next reviewer. Others surface much later, during an audit or when a customer complains, and those are usually the most damaging.

How to answer it well. Describe the blast radius if the agent produces a wrong answer. Who is affected, who notices and how quickly, and how easily can the error be corrected. Be concrete; the honesty of this answer protects you later.

> EXAMPLE ANSWER
>
> "A poor draft wastes the consultant's review time. At worst the consultant misses an inaccuracy that reaches the client, damaging the engagement but not committing the firm to anything irreversible."

How much detail is enough?

A strong set of answers is one that someone outside the project could read and understand exactly what the agent does, what it touches, what it delivers, and what an error would mean.

If the answers do not yet feel that clear

That is normal on a first pass. Take more time. Talk to the people who do this work today, or who would own it once the agent went in. Try the questions again with what you have learned.

The same questions across different industries

The questions stay the same. The answers vary by industry.

CONSULTING

A research briefing agent

An agent that drafts internal client-meeting briefings for consultants to review. Question 3 answer: "Producing information for the consultant." A natural fit for the lightest level of autonomy.

INSURANCE

A claims summarisation agent

An agent that reads a claim file and produces a summary with recommended next steps for the adjuster. Question 3 answer: "Recommending specific next steps." A heavier level, with careful checks on each fact extracted.

BANKING

A credit memo drafting agent

An agent that drafts credit memos for relationship managers preparing for credit committee. Question 3 answer: "Recommending specific positions." A heavier level, with first-line ownership and second-line review.

Classify your project against four levels.

AI agents differ in how much they do on their own. The level of autonomy shapes what controls the agent needs around it, who has to be involved when it runs, and how long it takes to bring into production.

There are four levels, from least autonomous to most. Your answer to Question 3, supported by Questions 5 and 6, points you toward the right level for your project.

Level 1. Provides information.	Produces information for a person to use. Example: a research briefing agent.
Level 2. Recommends actions.	Proposes specific actions for a person to take. Example: a contract review agent.
Level 3. Acts with human approval.	Performs actions, gated by a person's sign-off. Example: a claims agent.
Level 4. Acts independently within policy.	Acts on its own inside an enforced policy. Example: a fraud screening agent.

LEVEL 01 · LIGHTEST

Provides information

What the agent does. Produces information for a person to read and use. The person decides what to do and does it.

Example. A research briefing agent that drafts notes for an engagement team. The agent never reaches a client system. It produces text the consultant reads and uses.

A common first project. Most firms start here. The agent never reaches a customer or downstream system without a person in between.

LEVEL 02

Recommends actions

What the agent does. Recommends specific actions; the person still chooses whether to act on each one.

Example. A contract review agent that flags clauses and proposes revised wording for the lawyer to accept, edit, or reject.

A common second project. The output is consequential because it shapes what a person will do, but the person owns the action.

LEVEL 03

Acts with human approval

What the agent does. Does the work itself, but every action it proposes is gated by a human sign-off before it commits.

Example. A claims agent that drafts payout decisions and waits for the adjuster to approve. The agent does the bulk of the work; the person owns the decision per case.

Usually not the first project. Most firms reach this level on a second or third project, once a lighter agent has built the operational discipline it needs.

LEVEL 04 · HEAVIEST

Acts independently within policy

What the agent does. Acts on its own inside an enforced policy. A person reviews by exception, when the policy is breached or the agent flags uncertainty.

Example. A fraud screening agent that declines transactions automatically using audited rules.

Almost never the first project. The depth this level needs has to be earned with lighter agents first.

What each level needs to run well.

Each level needs a specific set of supports to run reliably. The table names them concretely. Read across the level your project sits at.

LEVEL	WHAT IT IS	EXAMPLE USE CASE	OWNERSHIP AND OVERSIGHT	RISKS TO MANAGE
Provides information	Produces information for a person to read and act on.	A research briefing agent that drafts internal client-meeting notes.	Sample-based quality checks. A reviewer always sits between the agent and the customer.	Inaccuracies the reviewer misses. Drift in output quality over time.
Recommends actions	Recommends specific actions. A person decides whether to act.	A contract review agent that flags clauses for a lawyer to review.	Recommendations logged. Quality reviewed against the actual decisions taken.	Reviewers rubber-stamping. Recommendations drifting from the firm's risk tolerance.
Acts with human approval	Performs actions. Each is gated by a person's sign-off before it commits.	A claims agent that drafts payouts for an adjuster to approve.	Every action logged. Code controls prevent acting without approval. A defined approver pool.	Approval bottlenecks at volume. Approvers missing the context to catch errors.
Acts independently within policy	Acts on its own within an enforced policy. A person reviews by exception only.	A fraud screening agent that declines transactions under audited rules.	Code controls enforce the policy. Monitoring of every action (e.g. API call logs). A named stop protocol.	Policy drift over time. Attempts to exploit policy edges. Errors at scale.

Worked example.

Two candidate projects, one decision.

A consulting firm has two AI agent projects on the table. The example shows how one is an attractive first choice and the other less so, even though its numbers look bigger.

PROJECT A

A research briefing agent

Reads internal knowledge and external research; drafts briefings for client meetings. A senior consultant currently spends 4 to 6 hours per briefing across 200 briefings per quarter.

Level of autonomy: Provides information.

Indicative value: around 1,000 senior-consultant hours per quarter freed for higher-value work.

PROJECT B

A contract review agent

Reads commercial contracts and drafts the recommendations the firm sends to clients. Currently a senior consultant and a lawyer, across roughly 50 engagements per quarter.

Level of autonomy: Recommends actions.

Indicative value: around 30% reduction in contract-review cycle time across 50 engagements.

THE RECOMMENDATION

Build Project A first. It is lighter to run, contained inside the firm, and recoverable if it fails. The firm proves the approach, builds confidence with AI agents, and develops the operational habits the heavier projects will need. Project B has a higher value ceiling and is a strong second project, once Project A is running well.

Three short next steps.

You now have a clear view of one possible AI agent project: what it would do, what level of autonomy it would sit at, and what would be needed to run it well.

01. Try the same questions on your other project ideas.

Each project will sit at a different level of autonomy and need a different set of supports. The strongest first choice is usually the one that delivers real value at a level your firm is ready to take on now.

02. Plan the wider journey.

Choosing the right first project is the first of six steps in the Bounded Agency approach. The later steps cover how the agent is designed, what controls it needs, how it fits into your operating model, and how it runs in production. Each is the subject of a separate guide in this series.

03. Talk to us.

Whichever project you have in mind, a conversation with Serpin is a useful next step. We help firms move from a strong first AI agent to a portfolio of agents running reliably in production. Bring the project, the answers, or the questions you are sitting with.

Have a conversation with Serpin.

Book a 30-minute call to talk about your project where it sits across the four levels of autonomy and what good would look like for you.

BOOK A CALL
serpin.ai/booking

- hello@serpin.ai
- +44 207 101 4147