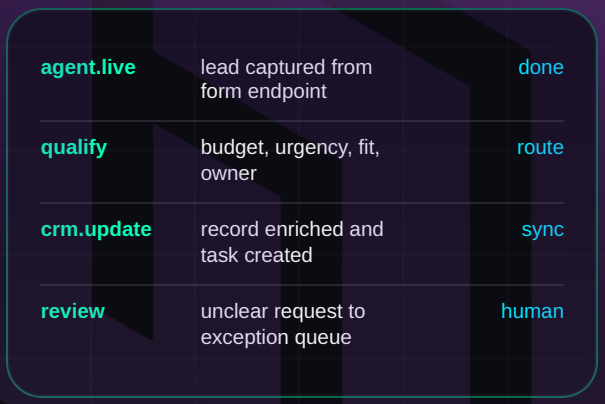
MARSHAL FIELD GUIDE / AGENTIC OPERATIONS



The CEO's Guide to Small Business AI Agents

Turn AI agents into operating leverage without buying another science fair project.

Kurt Fischman

CEO and Founder of Marshal

Table of contents

This guide is built for the CEO who has heard enough AI fairy tales and wants a working operating model. Miracles remain outside scope, as usual.

INTRODUCTION

From AI toys to operating leverage

The signal is noisy: many small businesses have touched AI. Far fewer have put it to work.

U.S. Chamber reports 58% of small businesses using generative AI. Census BTOS data saw overall business AI usage hovering between 17% and 20% in recent reporting windows.[1,2]

Most CEOs no longer need to be convinced that AI matters. You have seen the demos. You have watched your team use ChatGPT to draft emails, summarize calls, make slide outlines, rewrite job posts, and produce the occasional paragraph that sounds like it was raised by LinkedIn.

Useful? Sure. Strategic? Usually not. The difference between an AI toy and an AI agent is whether work actually moves forward inside the business.

A real agent has a job. It knows when to run. It uses the right company context. It takes allowed actions in your existing tools. It escalates when judgment matters. It leaves evidence behind. It improves over time. That last sentence is where most vendor demos die, quietly and without a memorial service.

The CEO question is not, "Can we use AI?" The useful question is: "Which recurring workflow can we safely transfer from manual effort to a managed system?"

THE CEO QUESTIONS THAT MATTER

- > What business result should this agent change?
- > What exact workflow is it responsible for?
- > What systems, data, and permissions does it need?
- > What actions are allowed without approval?
- > Who owns the agent after launch?
- > What metric proves it is earning its keep?

Without those answers, you get agent sprawl: ten pilots, five chat interfaces, a Slack channel full of screenshots, and no one brave enough to tell the CFO what changed. This guide gives you the operating model to avoid that circus.

OPERATING MODEL

The Small Business Agent Lifecycle

The Small Business Agent Lifecycle is a seven-step operating model for taking an agent from an idea to a production workflow. Its purpose is not to make a small company cosplay as a Fortune 500 transformation office. Humanity has suffered enough. Its purpose is to make sure every agent is tied to a real business outcome, tested against the work it will actually do, and managed after launch.

> McKinsey's 2025 State of AI survey found that 23% of respondents were scaling agentic AI somewhere in the enterprise and another 39% were experimenting, but no individual function had more than 10% of respondents scaling agents. The gap is the story: demos are easy; operating leverage is harder.[3]

01	02	03	04	05	06	07
Outcome Define the business result.	**Workflow** Map the trigger, handoffs, and output.	**Scorecard** Set metrics, thresholds, and rollback rules.	**Context** Connect data, tools, examples, and permissions.	**Build** Test, shadow, and harden.	**Launch** Pilot in the workflow.	**Run** Monitor, improve, and expand.

Gartner predicts more than 40% of agentic AI projects will be canceled by the end of 2027 because of escalating costs, unclear value, or inadequate risk controls.[4]

That prediction is not an indictment of agents. It is an indictment of sloppy adoption. Most failed agent projects are not too ambitious. They are too vague. The agent has no owner, no scorecard, no permissions model, no feedback loop, and no adult supervision. Then everyone acts surprised when the digital intern wanders into accounting with a flamethrower.

The seven steps in practice

Step 1: Outcome	Define the business case before anyone touches a builder, API key, or workflow canvas. What is the current process? What cost, delay, risk, or revenue leak does it create? What changes when the agent works?
Step 2: Workflow	Specify the unit of work. Name the trigger, inputs, decisions, allowed actions, handoffs, edge cases, and final output. If the workflow cannot be described in one clean paragraph, it is not ready to build.
Step 3: Scorecard	Choose a small set of metrics. Pair business KPIs with agent quality metrics and user feedback. Define promote, pause, and rollback thresholds before the pilot starts.
Step 4: Context	Assemble the minimum viable data, tools, permissions, examples, and feedback signals the agent needs. Use real constraints. Demo data is where governance goes to die wearing a tasteful blazer.
Step 5: Build	Build the workflow, then test it against golden examples, ugly edge cases, and real inputs. Run it in shadow mode before giving it authority in production.
Step 6: Launch	Roll out to a narrow pilot cohort. Train users on when to trust, edit, approve, or reject the agent. Make feedback easy and visible.
Step 7: Run	Monitor performance, exceptions, cost, and drift. Keep a runbook, a clear owner, and a way to throttle or stop the agent when reality decides to be annoying again.

What measurable ROI actually looks like

Boards, lenders, investors, and operators do not care that the agent used a fashionable model. They care whether the business now handles more work with the same team, responds faster, makes fewer mistakes, captures more revenue, or reduces operational risk.

The lifecycle makes ROI boring in the best possible way. You write the value narrative before the build. You set the scorecard before launch. You monitor the actual workflow after launch. No mystical spreadsheet ceremony required.

THE CREDIBLE ROI PATTERN

- Start with the before and after.
- Name the workflow and user population.
- Show the conservative time or cost math.
- Add observable business lift, not wishful vibes.
- Separate saved time from increased throughput.
- Track error reduction, exception volume, and customer experience.

bad ROI	"The agent saves everyone two hours per week." No one changed capacity, throughput, or spend.	flimsy
good ROI	"Inbound lead response fell from 11 hours to 6 minutes, booking rate rose 18%, and two weekly admin blocks disappeared."	useful
CEO lens	Time returned only matters when it converts into more output, lower cost, better speed, or revenue protection.	true

BCG has reported early adopters seeing 20% to 30% faster workflow cycles in ERP and CRM orchestration patterns. Treat that as a directional signal, not a promise. Your own workflow scorecard is the only number that can cash a check.[5]

EXAMPLE BUILD

Deploying a Speed-to-Lead and Intake Agent

Let us walk through a common first agent for founder-led businesses: a Speed-to-Lead and Client Intake Agent. It is not glamorous. That is a virtue. Glamour is what software vendors use to distract you while your best leads age like cheese in an inbox.

THE PROBLEM

Inbound leads arrive through website forms, email, ads, referrals, chat, phone notes, or some doomed spreadsheet named "new leads final v7." Response time varies. Qualification is inconsistent. Follow-up depends on who remembered to check the CRM. The agent's job is to detect new demand, qualify it against clear rules, route it to the right owner, draft the first response, book or propose times, update the CRM, and surface exceptions for human review.

Without the agent

- Leads wait hours or days for a human to notice.
- Owners qualify differently, so pipeline quality is foggy.
- Follow-up cadence depends on memory and caffeine.
- CRM records are incomplete, duplicated, or just spiritually disappointing.
- Sales calls are booked late or not at all.

With the agent

- Every lead is acknowledged quickly.
- Fit is scored against the same qualification rules.
- The right owner gets the right context.
- Draft replies and next steps are queued for review where needed.
- Response speed, booking rate, and stale lead rate are visible from day one.

Outcome, workflow, and scorecard

OUTCOME

People involved: CEO, sales owner, operations lead, and whoever currently handles inbound leads while pretending the process is "mostly fine."

Key deliverable: A one-page value narrative with the current response process, expected operating change, pilot scope, and KPI baselines.

WORKFLOW

Trigger: A new lead appears in a form, inbox, ad platform, CRM, chat, or referral workflow.

Output: Qualified lead record, routed owner, draft response or booking path, CRM update, and exception if the agent lacks confidence.

In scope

- Lead capture and source detection.
- ICP scoring and qualification.
- Routing by segment, service, geography, or owner.
- Draft email or SMS response for approval.
- Calendar availability checks and booking suggestions.
- CRM enrichment, task creation, and logging.

Out of scope

- Sending sensitive or high-stakes messages without review.
- Changing deal stages beyond allowed pilot rules.
- Accessing payment, legal, or private customer data unless explicitly approved.
- Quoting custom pricing without human sign-off.
- Handling leads outside pilot channels.

METRIC	WHAT IT MEASURES	DIRECTION
Lead response time	Time from inbound trigger to first useful response or routed action.	Decrease
Qualification completion	Percent of leads scored with enough information for a next step.	Increase
Booked meeting rate	Qualified leads that book a call or receive a clear scheduling path.	Increase
Stale lead rate	Qualified leads with no owner action after a defined SLA.	Decrease
CRM completeness	Required fields populated accurately after intake.	Increase

Context and build

Context is where most small business AI projects either become useful or become a very expensive hallucination machine with a friendly loading spinner. The agent must run on your real tools, your real rules, and your real permissions.

DATA SOURCES

- CRM records and lead source fields.
- Website form submissions and chat logs.
- Sales inbox or shared mailbox.
- Calendar availability.
- ICP rules, service definitions, territories, and pricing guardrails.
- Examples of good and bad lead handling.

ALLOWED ACTIONS

- Create or update lead records.
- Score and route leads.
- Draft replies for human review.
- Create tasks and Slack notifications.
- Suggest meeting times.
- Escalate unclear or risky cases to an exception queue.

TESTING PHASE	WHAT YOU ARE CHECKING	WHO REVIEWS
Golden lead set	Historical leads get the right score, route, and draft action.	Sales owner and ops lead
Messy edge cases	Vague, spammy, duplicate, urgent, and out-of-scope leads do not break the process.	CEO and functional owner
Shadow mode	The agent runs on live leads but actions stay draft-only or internal.	Design partner users
Pilot cohort	Acceptance, edit patterns, booking results, and exception rates are stable.	Program owner

> Do not skip shadow mode. The agent should encounter reality before reality encounters your customers.

Launch, monitor, and improve

Launch is not a Slack announcement with confetti emojis. Launch is a controlled change to how work happens. Start with one channel or one owner group, keep review gates tight, and expand only when the scorecard says the agent is earning more authority.

LAUNCH PLAN

- One inbound channel or lead type.
- One owner for pilot performance.
- Clear approval rules for outbound messages.
- Published feedback path for bad suggestions.
- Promotion criteria and rollback criteria agreed in advance.

RUNBOOK

- How to pause the agent.
- How to change routing rules.
- Who reviews exceptions daily.
- Who owns CRM field quality.
- What happens when the model or integration changes.

SIGNAL	WHAT IT TELLS YOU	ACTION IF DEGRADING
Draft acceptance rate	Whether humans trust the suggested response.	Review low-acceptance examples and tune instructions.
Exception rate	Whether the agent is over-scoped or missing context.	Narrow scope or add source data.
Booking completion	Whether qualified leads reach the calendar.	Inspect handoff, message quality, and availability rules.
CRM error rate	Whether automation is damaging system hygiene.	Pause writes, fix mapping, retest.
Stale lead rate	Whether owners are still dropping the ball after routing.	Escalate SLA alerts and owner accountability.

Why this example matters

Speed-to-lead is a good first agent because it is narrow, frequent, measurable, and tied to money. It sits at the edge of the business where missed work becomes missed revenue. That makes it a useful proving ground for the operating model.

LEAD CAPTURE

Capture, qualify, route, book, and follow up. The first place operational slop turns into revenue leakage.

REVENUE GENERATION

Research accounts, enrich contacts, draft outbound, and coordinate the next best action for sellers.

OPERATIONAL THROUGHPUT

Move intake, onboarding, admin relay, reporting, and internal handoffs through the business.

foundation	The same context, permissions, and feedback loop support the next workflow.	reuse
compound	Lead capture feeds sales execution. Sales feeds onboarding. Onboarding feeds reporting.	scale
lesson	The first agent is not a toy. It is the first brick in an operating layer.	armed

> The point is not to automate everything. It is to find the recurring work where speed, consistency, and follow-through matter more than heroic manual effort.

TEAM DESIGN

Building your operating model

A lifecycle on a slide is easy. A lifecycle that drives ROI requires people who own the decisions, the workflow, the data, and the post-launch reality. Small businesses do not need an AI council with laminated governance theater. They need a compact operating model.

ROLE	WHAT THEY OWN	COMMON MISTAKE
CEO or owner	Priority, business outcome, funding decision, and accountability for adoption.	Delegating strategy to whoever likes prompts.
Functional owner	The workflow, the scorecard, edge cases, and what "good" looks like.	Asking for a generic bot instead of a specific job.
Ops lead	Process mapping, handoffs, implementation sequence, and feedback cadence.	Letting tool configuration replace process design.
Technical owner or operator	Integrations, permissions, monitoring, and model/tool changes.	Shipping without observability.
Human reviewers	Approval, correction, quality feedback, and exception handling.	Rubber-stamping outputs until the agent learns the wrong lesson.
Risk owner	Data boundaries, customer impact, compliance, and escalation rules.	Bolting governance on after the first public mistake.

> NIST frames AI risk management as a way to incorporate trustworthiness into design, development, use, and evaluation. For SMBs, translate that into practical controls: permissions, approval gates, logging, rollback, and a human who knows where the bodies are buried.[6]

How to find the right first agents

Start with day-in-the-life mapping. Follow a lead, ticket, order, invoice, onboarding step, or weekly report from trigger to finish. Do not start with software. Start with the human choreography your business already runs and complains about in meetings.

- Where does work wait for someone to notice?
- Where is the same judgment applied over and over?
- Where do handoffs fail?
- Where does missing context create rework?
- Where does slow response cost revenue or trust?
- Where does a manager review work without changing much?

THE FIRST-AGENT FILTER

High frequency	Runs often enough to learn from real volume.
Clear trigger	Starts from an observable event.
Known output	Produces a record, draft, route, report, or task.
Low catastrophic risk	Mistakes are recoverable under review.
Visible metric	Business impact can be measured in days or weeks.

1

Pick the workflow

One trigger, one output, one owner, one measurable result.

2

Run the lifecycle

Outcome, workflow, scorecard, context, build, launch, run.

3

Make the proof portable

The first production agent becomes your internal standard for the next one.

If you cannot define the trigger, allowed action, and success metric in an hour, you are not ready to build. You are ready to argue in a conference room, which is cheaper but less useful.

INFRASTRUCTURE

What makes SMB agents scale: Managed AI Ops

Small businesses do not need an enterprise platform committee. They do need a shared operating layer. Otherwise every agent gets its own brittle retrieval setup, its own integration hacks, its own permission assumptions, and its own little pile of invisible risk. That is not leverage. That is a junk drawer with API keys.

1. Unified business context

Agents need access to the policies, docs, CRM records, emails, tasks, customers, and prior decisions that define how the business works.

2. Governed tool use

Actions must run through scoped permissions, approval gates, and exception queues so the agent can move work without freelancing.

3. Observability

You need visibility into what the agent did, why it did it, where it got stuck, and what it cost.

4. Managed improvement

Agents drift when data, tools, rules, people, and models change. Someone has to operate the system after launch.

tools	HubSpot, Gmail, Calendar, Slack, Stripe, Notion, Airtable	connect
context	Docs, records, owners, policies, history, customer notes	ground
controls	Permissions, approval gates, logs, exception queues	govern
agents	Lead capture, revenue generation, operational throughput	execute
ops	Monitor, tune, report, expand, kill when needed	run

Managed AI Ops is the practical answer for founder-led companies that want production agents without hiring an internal agent engineering team. The system runs on the stack you already own, with humans in command where judgment matters.

What to look for before you buy or build

Do not evaluate agents by the demo. Every demo is trained to behave for five minutes under studio lighting. Evaluate the system by how it behaves after launch, when the CRM schema changed, the owner is out sick, a customer writes something weird, and the API limit gets moody.

> OWASP names prompt injection, sensitive information disclosure, excessive agency, and overreliance among key LLM application risks. For agents, those are not abstract security trivia. They are reasons to limit tools, validate outputs, log actions, and keep humans in the loop.[7]

PLATFORM AND OPERATOR CHECKLIST

- Connects to existing tools without forcing migration.
- Uses permission-aware context, not a loose pile of uploaded files.
- Supports approval gates for sensitive actions.
- Maintains exception queues and human review paths.
- Logs inputs, actions, sources, and outcomes.
- Has evaluation metrics for quality and business impact.
- Includes monitoring, runbooks, and rollback.
- Can expand from one workflow to a portfolio.

QUESTION	WHAT A GOOD ANSWER SOUNDS LIKE
What does the agent own?	A named workflow with a trigger, allowed actions, output, and owner.
What can it touch?	Only approved systems, fields, records, and actions.
How does it fail safely?	It escalates, drafts, pauses, or routes to a human before damage spreads.
How do we know it works?	The scorecard tracks business KPIs, quality signals, and user feedback.
Who fixes it in month six?	A named operator, not a support queue praying for closure.

MEASUREMENT

Measuring, communicating, and funding agent ROI

Because you ran the lifecycle, you are not inventing ROI after the fact. You have the outcome narrative, scorecard, pilot data, user feedback, exception log, and improvement history. In other words, you have evidence. Evidence is what survives meetings after enthusiasm leaves the room.

A SIMPLE PER-AGENT ROI FRAME

Agent ROI = time returned + throughput gained + revenue captured + risk reduced - operating cost.

Time returned is the weakest input unless it changes capacity or cost. Throughput, revenue capture, quality, and risk reduction usually tell the stronger story.

Productivity agents

Return time to owners and staff by removing repeated admin, drafting, search, summarization, and coordination work.

Throughput agents

Increase volume, consistency, or speed in intake, support, onboarding, reporting, and delivery.

Revenue and risk agents

Protect or generate dollars by improving response speed, pipeline execution, collections, QA, compliance, and escalation.

SPEED-TO-LEAD ROI INPUT	EXAMPLE CALCULATION
Manual time removed	80 leads per month x 15 minutes x $60/hour = $1,200/month of labor capacity.
Throughput gained	Every qualified lead gets routed and followed up inside the SLA, not whenever someone remembers.
Revenue captured	One additional qualified meeting or customer can outweigh the labor savings by an embarrassing margin.
Risk reduced	Fewer stale leads, fewer missed follow-ups, cleaner records, and a visible exception queue.

Using ROI to guide funding and roadmap decisions

Once every agent uses the same lifecycle, roadmap decisions get much less theatrical. You stop funding whoever made the prettiest demo and start funding the workflows that prove leverage.

FUNDING RULES

- **Kill** agents that do not clear the scorecard.
- **Keep local** agents that help one team but do not justify expansion.
- **Scale** agents with stable quality and visible business lift.
- **Manage like a product** agents that become critical workflow infrastructure.

VENDOR QUESTIONS

- Which part of the lifecycle do you own?
- What happens when the agent fails?
- Can we keep the data, workflows, and assets if we leave?
- How do you monitor production behavior?
- What gets human review and why?

TIMEFRAME	CEO FOCUS	EXPECTED OUTPUT
0 to 30 days	Pick one high-frequency workflow and define the outcome, workflow, scorecard, and context.	One scoped pilot ready to build.
30 to 60 days	Build, test, shadow, and launch with narrow authority.	One working agent in a real workflow.
60 to 90 days	Measure performance, tune behavior, decide whether to scale, narrow, or kill.	A repeatable proof pattern.
3 to 6 months	Expand to related workflows using the same context and controls.	An agent portfolio with operating cadence.
6 to 12 months	Treat critical agents as managed operating infrastructure.	AI leverage that compounds across the business.

> Do not spin up ten agents before one survives production. The first real agent is not just a use case. It is your proof of how the business will operationalize AI without becoming a vendor-funded petting zoo.

CONCLUSION

Bringing it all together

AI agents are going to show up inside your business whether you plan for them or not. They will arrive through chat tools, CRM features, inbox copilots, workflow platforms, marketing tools, customer support apps, and the enthusiastic employee who discovers a new product at midnight and returns Monday with the confidence of a minor prophet.

The CEO's job is not to stop the future at the door. That would be adorable. The job is to decide whether agents become a coherent operating layer or a mess of disconnected experiments.

The winners will not be the companies with the most prompts. They will be the companies with the clearest workflows, strongest context, safest permissions, tightest feedback loops, and most disciplined measurement.

Start with one painful workflow. Run the lifecycle. Build the evidence. Expand only after the first agent earns more responsibility. That is how AI becomes leverage instead of decoration.

Find your first agent at runmarshal.com/consult

Marshaling intelligence for businesses that play to win.

Marshal builds, deploys, and manages AI agent systems for founder-led businesses.

REFERENCES

Sources and notes

1. U.S. Chamber of Commerce, "The Majority of Small Businesses Embrace Artificial Intelligence," 2025. https://www.uschamber.com/technology/empowering-small-business-the-impact-of-technology-on-u-s-small-business
2. U.S. Census Bureau, "AI Use at U.S. Businesses," 2026. https://www.census.gov/library/stories/2026/05/ai-use-businesses.html
3. McKinsey, "The state of AI in 2025: Agents, innovation, and transformation," 2025. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai
4. Gartner, "Gartner Predicts Over 40% of Agentic AI Projects Will Be Canceled by End of 2027," 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
5. Boston Consulting Group, "How Agentic AI is Transforming Enterprise Platforms," 2025. https://www.bcg.com/publications/2025/how-agentic-ai-is-transforming-enterprise-platforms
6. National Institute of Standards and Technology, "AI Risk Management Framework," overview and AI RMF 1.0. https://www.nist.gov/itl/ai-risk-management-framework
7. OWASP Foundation, "Top 10 for Large Language Model Applications." https://owasp.org/www-project-top-10-for-large-language-model-applications/
8. Marshal, company and platform materials, including Managed AI Ops, AI Agent Systems, Agent Governance, and Marshal Agents. https://www.runmarshal.com

ABOUT MARSHAL

Marshal is a Managed AI Ops service for founder-led businesses. We design, build, deploy, and manage AI agents that take recurring work off the team and help businesses get found inside answer engines.