

— A SOCRATIX POINT OF VIEW

The Future of Data and Organizations in the AI Era

Your data is a mess. That stopped being the problem.

HUMAN WISDOM. AI ILLUMINATION.

Socratix AI • Cleveland, Ohio • October 2026
socratixai.com • info@socratixai.com

AI in the building. Almost none of it in the business.

86%

of middle-market executives say AI is integrated into their operations

RSM · Middle Market AI Survey 2026 · n = 1,030

36%

have it embedded in core processes

RSM · 2026

34%

name data quality as the top barrier. Legacy integration is next at 28%.

RSM · 2026

16%

of mid-market companies have a fully governed, integrated data state

Kaufman Rossin · State of AI in the Mid-Market, 2026

Among manufacturers, 55% named legacy integration as the top barrier. Not one in Kaufman Rossin's sample had reached company-wide deployment.

Most mid-market companies have AI in the building and almost none of them have it in the business. RSM surveyed 1,030 middle-market executives in the U.S. and Canada this year and 86% said AI was integrated into their operations. Only 36% had it embedded in core processes. When RSM asked what was in the way, the answer wasn't cost and it wasn't talent. Data quality came first at 34%, then legacy systems integration at 28%. Kaufman Rossin's mid-market study from the same summer is blunter: 16% of mid-market companies have what they'd call a fully governed, integrated data state, and in manufacturing 55% named legacy integration as the top barrier. Not one manufacturer in that sample had gotten to company-wide deployment.

We run into this in almost every discovery call. The pilot worked, somebody built a prompt that saves them an hour a day, and then the company tries something that touches the ERP or the quoting spreadsheet or the twelve years of PDFs sitting in a shared drive, and it stops. The diagnosis is almost always some version of "our data is a mess, we have to fix the data first."

That diagnosis is half right. Your data is a mess. So is everyone's. Order history is in the ERP, pricing is in a spreadsheet one person maintains, the customer context is in email and the reason a job got expedited last Tuesday is in somebody's head. That's what a normal mid-market company looks like.

Where it goes wrong is the prescription. The two-year ERP replacement, the data warehouse project that has to finish before any AI can start. That advice comes from a world where software could only read data that had been structured for it, and that world is gone. Companies waiting on a clean-data milestone are waiting for something they don't need anymore, and they're spending the AI budget on plumbing while the shop down the road ships.

Software can now read the mess.

For forty years the rule was that you structured the data first and then the computer could use it. Every field in the ERP, every column in the spreadsheet, every validation rule is there because the machine couldn't understand anything else, and data entry was the job of translating reality into something a system could hold.

That's the rule that broke. A language model will read a work order with handwritten notes in the margin. It will read the email thread where the customer changed the delivery date twice, the spreadsheet where column C means one thing through row 40 and something else after that, your ERP export next to your vendor's PDF, and reconcile the two. It doesn't need the data clean. It needs to be able to get at it.

Some of what we've seen. At one manufacturer we work with, incoming work orders used to be reviewed and keyed in by hand. An AI review step now reads each order, checks it, and flags what's off before anything gets entered. That took out 75-plus hours a year of manual data entry and caught 40-plus input errors a year before they turned into wrong parts.

For another client we built an agent that reads employee calendars once a month and works out who took vacation. Nobody cleaned up the calendars first. One person writes "gV" to mean nine hours of vacation. The agent reads it that way, lists what it found, and each employee confirms or corrects it before a manager signs off and the record gets written. I keep coming back to the gV thing because it's the whole argument in two characters. Nobody was ever going to standardize that calendar. They didn't have to.

And a colleague of ours knows a business administrator at a government agency, no technical training, who is in practice the entire data function for his organization. Spreadsheets come in, he runs them through Copilot, the analysis goes back out, and his coworkers think he's a wizard. The capability is already in the hands of people with no data skills at all. Most organizations haven't noticed.

So the cost of making data usable collapsed. What used to be a six-month integration project is a prompt and a permission. The pile of PDFs, the inconsistent spreadsheet, the ERP nobody wants to touch stopped being blockers and became inputs.

Which is not the same as saying there's no risk in it. The agent that reads gV as nine hours is going to be right nearly every time, and the entire design question is what happens the other times. That turns out to be an organizational question more than a technical one.

HOW THE CALENDAR AGENT WORKS

Agent reads

the shared calendar once a month, informal notes included

Employee confirms

a card listing what it found, with hours to edit

Manager approves

the confirmed list, or sends it back

Written to the record

through the same door a person's entry would use

A simplified version of a real build. Accounting can still open the record and change it by hand.

Three things change.

The first thing it changes is what "mature" data even means. The old scorecard asked whether your data was standardized, deduplicated, governed. The question that matters now is whether an agent can reach it with the right permissions, and a spotless warehouse nobody has wired an agent into is worth less than a messy SharePoint folder with access scoped properly. Clean what the agents touch, when they touch it. Spend the rest of the cleanup budget on access.

The second is where the data that actually runs the business lives. The ERP holds transactions. The decisions live somewhere else: why the order was expedited, what the customer actually agreed to on the phone, which vendor gets the call when the usual one is out. All of that sits in email, in Teams, in text messages, in meeting notes and in people's heads, and it was never called data because nothing could use it. Now something can. The more of these projects we do the more it looks like this is where most of the value is, because it's the context that lets an agent handle an exception the way a twenty-year employee would. It's also the data most likely to walk out the door when that employee retires.

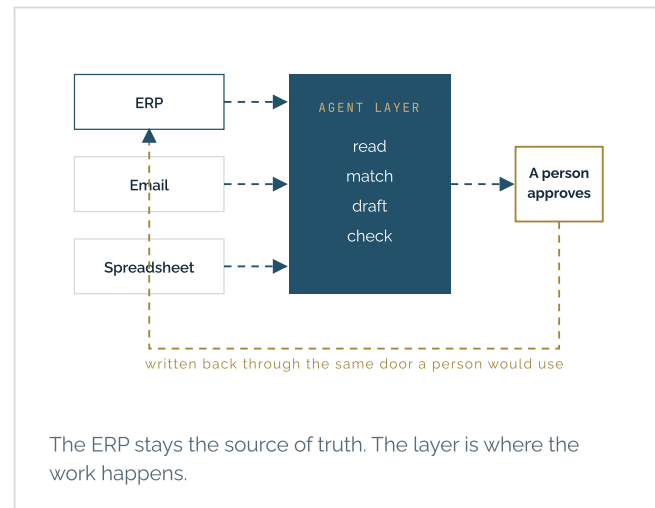
The third one is further out, but I'd rather say it now than in two years. Most of your data is about to be produced by agents. Every run leaves a trail: what it read, what it inferred, what it proposed, who approved it. Twenty agents will generate more operational records in a quarter than your people did in a year. That's good, you finally have an audit trail, and it's a new problem, because somebody has to decide what to keep, who checks it, and which system is the record. Treat agent output the way you'd treat a junior employee's work. Reviewed, logged, written to one place. Skip that and you've got twenty agents each keeping their own version of the truth.

No. Keep the system of record. Put the work in a thin layer around it.

That's the question we get most often, and the answer is no. For most companies an ERP replacement is the most expensive road to the least amount of AI. The ERP is your system of record. It was never meant to be where the work happens, and it isn't going to become that because the vendor added an AI tab.

Keep the ERP as the source of truth and put a thin layer around it where the work actually gets done. Agents read from the ERP, from email, from the spreadsheet. They do the reading and the matching and the drafting and the checking, a person approves, and the result goes back into the ERP through the same door a person would have used. The ERP never has to understand AI. It has to accept a well-formed record, which it already does every day.

What the agent needs from that layer (a way to read, a permission, a way to write back) is a paper of its own: *UI for AI, UI for Agents*.



Three shapes for the layer

Where that layer lives is mostly a question of how much control you want. For a mid-market company there are really three shapes, and since most of our clients are on Microsoft that's where the examples are.

Governed assistant

Microsoft 365 Copilot and Copilot Studio inside the tools you already own, inheriting whatever permissions and information-protection policies IT has already set. Fastest, cheapest, and the agent can only see what the user already could. That's where most first use cases belong.

CONTROL ●●○○○ · COST ●○○○○ · UPKEEP ●○○○○

Custom app on managed models

Your own application running on Azure-hosted models inside your tenant, which gets you more control over the workflow and more capability and more to maintain. You go there when the first shape runs out of room.

CONTROL ●●●○○ · COST ●●●○○ · UPKEEP ●●●○○

Model you operate yourself

Self-hosted, on compute you run. Maximum control, maximum cost, and almost never where a mid-market company should start.

CONTROL ●●●●● · COST ●●●●● · UPKEEP ●●●●●

One thing to say about all three. Security depends on the whole system, meaning who can retrieve what, where it gets processed, what's logged, who maintains the controls. "It's Microsoft, so it's fine" isn't an answer, and neither is "we self-host, so it's fine." Self-hosting doesn't make you safe. It makes you the one responsible.

Throw out the warehouse scorecard. Ask three questions, one process at a time.

Can an agent reach it?

Where does the data for this process live, and does anything other than a person have permission to read it? If the answer is a folder on someone's desktop, or "the ERP, and one person has the login," that's the first fix, and nine times out of ten it's a permissions change rather than a project.

Does a human know what good looks like?

The agent can read the order. Someone has to know this customer always means net-30 even when the PO says net-15, or that gV means nine hours. That judgment is the real asset and it belongs to the people who run the process. IT doesn't have it and shouldn't be expected to. Write it down and put it somewhere the agent can read it. For most teams that's the entire knowledge-management project and it takes an afternoon.

Where does the result get written back?

Every agent output needs a home the business already trusts. If the answer is a new spreadsheet the agent keeps for itself, stop. If it's the ERP, through the same approval a person's entry would get, go.

Two things fall out of asking it this way. Data stewardship moves from IT to operations: IT owns access and security, and the department that generates the data owns whether it's right and owns the rules the agent follows. That's a real shift in responsibility and it has to be said out loud, because most operations leaders have spent twenty years being told data is IT's problem.

WORKSHEET • SCORE ONE PROCESS

Process

Can an agent reach the data today? YES ☐ NO ☐ NOT SURE ☐

Does one person know what good looks like, and is it written down? YES ☐ NO ☐ NOT SURE ☐

Is there one trusted place the result would be written back to? YES ☐ NO ☐ NOT SURE ☐

Three yeses: build the smallest agent that reads it this week.
A no: that's the first fix, in this order. Permission, then the written-down rule, then the system of record.

And the data problem turns out to be partly a people problem. RSM found 85% of middle-market executives say leadership is more excited about AI than their employees are. The people who know what gV means are the people who decide whether the agent gets corrected or quietly worked around, so build it with them. Once somebody sees an agent reading their notes they start writing the notes so the agent gets them right, and that's cleanup you never had to fund.

WHERE TO START

One messy process where the pain is weekly.

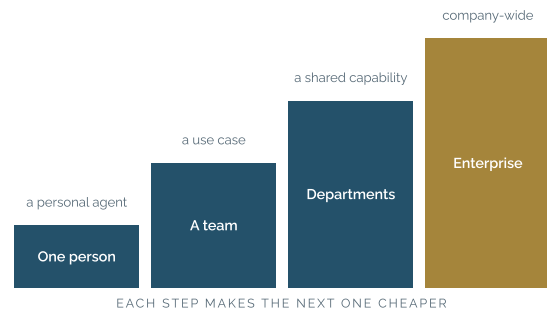
Not the biggest one. The one where somebody re-keys information from one system into another every Monday morning, or chases people for something that's already on a calendar, or reconciles two documents that should have matched in the first place.

Then ask the three questions about it, fix the permission, write down the rule, name the system of record. Build the smallest thing that reads it: an agent that takes the inputs, proposes the output and hands it to a person to approve. No autonomy on day one, because the approval step is how you learn the error rate and how the team learns to trust it. Measure hours and errors and nothing else, since those are the two numbers a CFO believes, and that manufacturer's 75 hours and 40 errors came out of one process. How to build the measurement so it holds up when budgets tighten is the whole of *Measuring What Matters*.

After that it widens on its own. A personal agent for one person becomes a use case for a team, then a shared capability across departments, and eventually something enterprise-wide. That's the path we're on with a port authority and with that manufacturer, and neither of them started with a data project.

The ERP replacement can wait. By the time you're ready for it you'll know exactly what you need it to do, because your agents will have shown you.

HOW IT WIDENS



KEEP READING

UI for AI, UI for Agents

Is your business ready to talk to a lot of agents? The other half of this paper, written from the agent's side.

socratixai.com

Measuring What Matters

A foundations-based approach to AI ROI. The bubble won't punish spending. It will punish unmeasurable spending.

socratixai.com

The Three-Foundation Framework

How mid-market companies turn AI from a curiosity into compounding value. The model behind everything we write.

socratixai.com

About Socratix AI

Socratix AI is a Cleveland firm that turns AI ambition into completed work for mid-market companies: strategy, training and hands-on implementation. We work across three foundations, People & Culture, Business Strategy, and Capability & Technology, because in our experience the data problem is never just a technology problem.

If you're stuck at the pilot wall and the advice you're getting is "fix the data first," we'd like to see the process. It starts with a conversation and a walk through the operation. info@socratixai.com · socratixai.com

Human Wisdom. AI Illumination.

SOURCES RSM Middle Market AI Survey 2026: U.S. and Canada (RSM US, July 2026) · Kaufman Rossin, The State of AI in the Mid-Market, as reported by MarketScale and Automation World (July 2026) · Client results from Socratix AI engagements, 2026.