Building New Enterprise Capabilities & Value

A Practical AI Playbook for Mid-Market Companies

ROTATIONAL LABS

Second Edition
September 2025

Executive Summary

This AI playbook is written for the **curious**, **bold**, and **forward-thinking mid-market executive, like Ted, whose journey we follow.** If that's not you, **stop reading now**.

Who are we talking about? You've led a mid-market enterprise for the past 8 years. You've got a lot on your plate. Complexity piles up; the board and customers are demanding. Vendors are raising prices; your IT infra is dated. New competitors emerge requiring you to cut prices. Tariffs have upended supply chains. It's exhausting.

We understand you may not have time to read a well-written, conscientious AI playbook grounded in real-world experiences. You have priorities.

So if you stop now, take this one idea with you: Disciplined Use. It's how you've succeeded as an organization; it's the key to unlocking the potential in AI now.

And then there's AI. You've been pitched numerous AI point solutions promising to solve all your problems. And what's your experience? Microsoft's AI Co-Pilot helps you write better emails: cool, but not ground-breaking. Or maybe you have tried several AI Proofs-of-Concept (POC), only to get stuck in "POC purgatory."

Yet, you are curious, bold, and forward-thinking enough to continue reading because you have a sense that **there is real and significant potential with AI**. From our vantage point as AI engineers with a collective 50 years of experience building enterprise AI solutions, we have three conspicuously non-AI principles to share with you:

- AI will not solve all your problems, but it does offer **an opportunity to re-define your business and build new capabilities**. You can build and deploy affordable, trusted, purpose-built AI agents *today* that do real work at scale, opening up new opportunities for your organization. **Your "design space" to automate and innovate across your org is vastly larger than ever before, if you have the vision**.
- AI must be **rooted in business fundamentals**. It's **a tool** that can be **combined with your hard-earned domain expertise** and applied to many valuable use cases across a variety of business functions such as marketing, operations, business intelligence, research, and finance. There is **so much low-hanging fruit, if you're genuinely curious**.
- Adapting AI so it does real work for your business is **not about tech, models, or data**, but **business alignment**, **collaboration**, **disciplined use and accountability**. That's how you build trust in AI. As a business operator, you know trust is the real currency, while process, discipline, and accountability lead to efficiency and value. *You know this already.*

We seek to bring these principles alive while recognizing the budget and time constraints of mid-market companies. If you continue to read, you'll learn:

- How new enterprise AI capabilities are built and deployed with disciplined use
- How to practice disciplined experimentation and use
- Practical insights into implementing AI agents that you won't see elsewhere

Note: This AI playbook is designed to be skimmed by execs. We know you're busy.

Table of Contents

Setting the Scene

The widespread adoption of generative AI techniques has thrown the knowledge economy into a period of profound transformation. This condensed AI playbook is for the experienced mid-market executive left wondering, "but what should *we* be doing with it?"

It's hard to get a straight answer. Data scientists focus on the algorithms and models at the expense of business value. Engineers focus on scalability but lack product context. Meanwhile executives are inundated with hucksters selling tools destined to become another contract to renew, renegotiate, or cancel at the end of the year, with no clear sense of how (or even *if*) it improves the bottom line.

We're not going to sugarcoat this. If you're picturing an IKEA-like catalog of AI solutions that will automagically yield ROI for your organization's unique combination of technology, people, and processes, you're thinking about it wrong. Automation is not firing magic bullets at clearly-marked targets. It's acknowledging dysfunction - and opportunity - in your operations and taking **a disciplined approach** to address it.

Resist the urge to simply imitate what other companies are doing and instead, **dive deep into the business processes at your organization**, however mundane they seem. You'll need to be genuinely curious, ask "why" questions, and be ok with dead-ends or initial failure to figure out what works at your organization. Learn to ask the right questions whose answers reveal the processes that *technically* work, but that are too slow, expensive, brittle, or opaque to scale. This is where agentic AI shines. **Your mission is to question - and ultimately change - the status quo.**

While there is no one-size-fits-all template for foolproof automation, **there are many best practices**, and we will share ours with you throughout this playbook (look for the Rotational lighthouse). We hope these tips will serve as a beacon for you on your journey.

First, the basics: What is AI good at?

Many business problems present as complex, noisy, and unstructured data: overwhelming for humans, but ideal for LLMs to interpret. AI agents can distill massive, fragmented information streams (reports, transactions, logs, documents) into clear, actionable insights or take action. We break AI's strengths into **four pillars**:

- **Understand** - summarize, interpret, translate, navigate knowledge
- **Decide** - pattern recognition, judgment, decision support
- **Create** - generate text, code, designs, simulations
- **Act** - execute workflows, automate processes, monitor results

Consider how much time and effort in your organization is dedicated to these activities. We've distilled these strengths into **discrete business tasks** amenable to AI agents. These tasks can be combined - or chained together - to build AI agents that automate workflows that deliver value. Read on to learn more.

Defining AI Agents and Co-Intelligence

When we talk about 'intelligent agents,' we don't mean artificial general intelligence (aka AGI), which is far from proven. We mean simple, well-scoped AI solutions: small language models and lightweight tools designed to help your team solve specific, practical everyday tasks. **We define an AI agent as a system consisting of a mission, instructions (prompts), a "brain" (LLM), tools, guardrails, and an environment**.

The combination of human talent and AI agents is what Ethan Mollick, an Associate Professor of Management at the Wharton School of Business, calls "co-intelligence." The goal of your AI automation journey is to **transform your organization into a co-intelligent firm**.

Meet Ted: Seasoned Executive vs Hype Machine

Ted (not his real name) is a late-50s exec at a 500-person IT company that does about $75 million a year in cloud support services for regional banks and hospitals. The brand is all about reliability and keeping the human touch in technology services, and they have a strong customer base.

Ted cut his teeth in the military, where he made a name for himself modernizing government data infrastructure, finally moving his agency from on-prem servers to the cloud. But later, after going to the private sector, he got burned by the big data craze. He had to learn the hard way that compute costs come with many promises, but often no clear return.

So when AI was suddenly everywhere, Ted knew better than to fall for the hype. He also understood that **his organization couldn't defend their market position if they didn't adapt.**

We follow Ted's journey to **identify an AI automation use case and build his organization's first agent**, which today reliably accelerates a core business process that once took several weeks and a team of five. Now it can be done in a single day, and only requires one human to act as supervisor. He laid a path for other departments in his organization to follow suit, and now they have a small army of agents supporting several functions quietly in the background.

As you'll see, the journey isn't linear. Ted's journey illustrates what it looks like to figure out what the *real* problem is, to clarify the AI use case, to build experiments and learn, and finally, to translate results into actions with measurable outcomes. It's not an overnight magic trick, but if you want this kind of transformation, here's how to begin.

Let's Talk About "Quick Wins"

Consumer AI like ChatGPT and Enterprise AI use LLMs, but **everything else is different**, and it's "what's different" that matters. Consumer AI companies spend millions monthly to hide complexity for low-stakes, individual tasks. Enterprise AI must tackle complexity: workflows, teams, high-stakes outcomes, messy data, and strict security. Only then does it unlock business value.

There is no "magic" or shortcuts in enterprise AI.

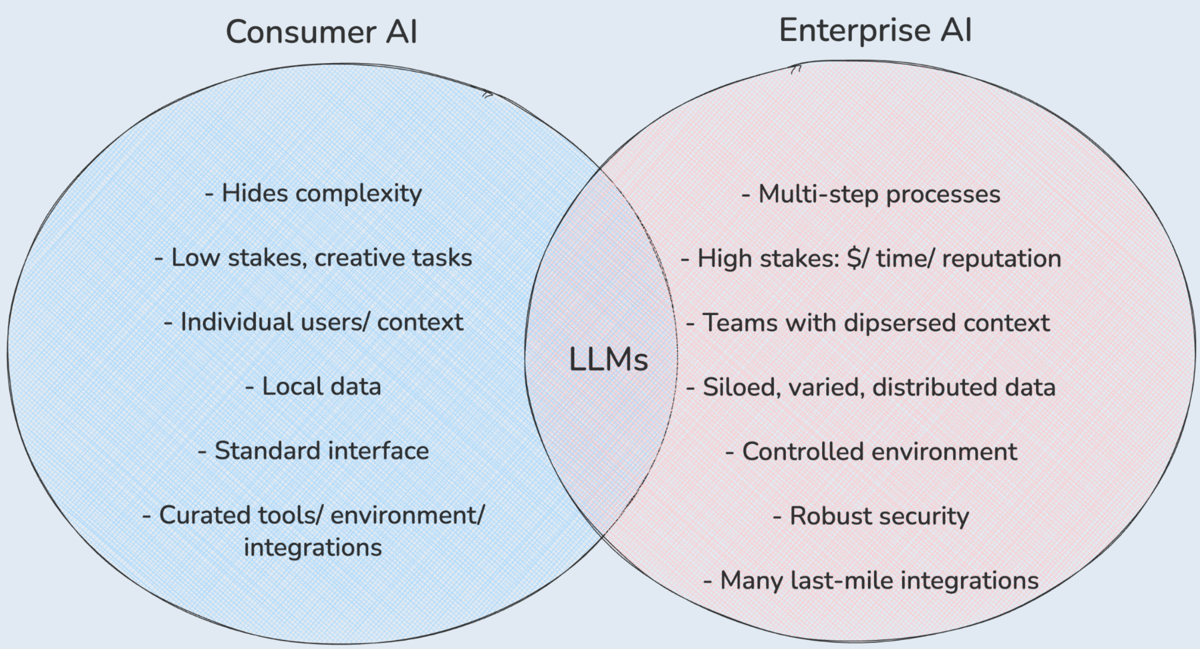

Figure Out What the Problem Actually Is

Ted's company had recently rolled out enterprise licenses for a new chat-based AI assistant and the board was eager to see use cases. But his instinct told him that the real problem was in how the team, processes, and tools fit together. Ted knew from experience on and off the battlefield that dysfunction isn't fixed by buying things. (If only life were so simple!)

Ted had an idea of where to start: a data analysis process that was a constant source of stress. It had started out as a research effort, but quickly gained traction to become a pillar of the company's revenue operations (revOps) strategy. Ted was fuzzy on the details, but it had something to do with taking in a lot of raw data and doing a series of complex analytic steps to produce a business intelligence report for customers.

When this process worked, it worked well. But two years in, it still took five people two weeks to make a single report. What worried Ted was not that it was slow, but that when it failed, it failed spectacularly. Ted winced whenever he thought about all the emergency Friday afternoon meetings he had been called into to discuss the latest analytical fiasco.

Ted realized he needed to get down on the field and develop a better understanding of the details. He began by talking with people on the team responsible for developing and maintaining the workflow. He discovered that while individually each member added value, together, they created friction:

- The **business analysts**, veteran consultants from the healthcare and financial sectors, really knew the customer; they deeply understood the data, and why it mattered. But their limited technical skills meant they often needed help with big or complex queries.

- The **data scientists** were the ones who had prototyped the new process, training several bespoke models that gave Ted's company a big leg up on the competition. However, they wrote "research code" and their work was hard to reproduce or audit when things went wrong (and it often did).

- To improve speed and reliability, several **engineers** had joined to help automate the process. These engineers were experienced coders, and some had even used AI tools, but they were not domain experts, so weren't always aligned with the rest of the team.

Finding Use Cases Through 3 Lenses: Ask the Right Questions

Focus on **use cases critical to business outcomes** that ultimately impact your bottom line. The real question isn't "What can we do with genAI?", but "What can genAI do for our business?"

Here are a set of questions to identify AI agent use cases through three business lenses.

Efficiency & Risk	People & Knowledge	Growth & Opportunity
• Which **processes take too long** or involve too many people? • Where do **mistakes** happen most often? • Where are **operating costs rising fastest**? • Which tasks are **highly repetitive** but still critical? • Where do **compliance** or audit risks show up?	• Where is **workforce turnove**r the highest? • What **expertise is concentrated** in just a few people? • Where do **teams waste time searching** for information? • What **business insights am I missing**?	• Where could **faster decision**s drive more revenue? • What would i**mprove the customer experience** most? • Where could we **scale a high-performing team** or workflow? • What could we **offer customers** if we had more capacity? • How could we **amplify our domain expertise**?

Critically, your AI use case should have **a business-aligned objective** grounded in economic reality.

An AI agent with a clear, business-aligned objective isn't just automation - it's a workforce multiplier that augments teams and drives outcomes tied to efficiency or growth. If the agent can't show impact on margins, it won't matter. If it can, you have a clear signal.

Early Signal & False Starts

The engineers were experimenting with the new chatbot, which could even help them code. They told Ted they thought they could build a minimal workflow that automated the entire analytic process end-to-end, no business analysts or data scientists needed. The idea was simple: feed the chatbot a massive prompt using the latest ultraXL GPT model. The prompt described the workflow to the best of their understanding, even embedding some key SQL queries from the data science team. Ted was a little skeptical, but he knew the board was eager to find use cases for the new enterprise licenses. "Go for it," he told them.

At first the proof-of-concept looked promising. In the demos, the prompt produced neat, professional-looking reports. But when the business analysts tried it for a real customer report, the model made strange assumptions and hallucinated. It was no more reliable than the humans it was supposed to replace. Ted thought it might actually be worse, since it was so confident about its blatantly wrong answers.

Beyond Chatbots

While they're an easy place to start, chatbots are rarely the path to solving dysfunction in the organization. For one thing, chatbots are not domain experts. Out of the box, they don't speak your language and are ignorant of the vocabularies, acronyms, and other domain-specific slang at your organization. Techniques like domain adaptation, alignment tuning, and retrieval augmented generation (RAG) can help correct for these deficiencies.

The engineers were optimistic, asking for a few more months to iron out the kinks. Ted wasn't so sure. It was starting to become clear that the real problem was that every client received their own *slightly customized* version of the report. What looked like one process from the outside was actually dozens of sovereign processes, each requiring context that only a domain expert would know.

The next day Ted got worse news. Their most experienced business analyst, the one who had quietly held the process together for two years, had just submitted her resignation. She had grown bored with the repetitive work and been poached by a competitor.

Boredom is a Sign of a Good AI Use Case

Because language models are good at extracting value from unstructured text, they can be very helpful for automating tasks that require a person to write, review, or evaluate a lot of documents. The telltale signs of these use cases are: employee turnover, customer churn, repetitive review tasks, tedious data entry duties, long onboarding processes, etc. These signals of organizational dysfunction are places where intelligent agents can help **reduce friction by automating the boring stuff**.

Ted could see the writing on the wall: even the best AI model on the market couldn't fill the gap left by human expertise, and when that expertise walked out the door in two weeks, it would be gone forever. To navigate the transition, they needed to capture and scale the knowledge that made the process valuable in the first place.

Now Let's Focus: Dive Deep into the AI Use Case

Ted knew the business analyst's last two weeks couldn't just be business as usual. She was the only domain expert who could tell the difference between a good report and a mess of AI slop.

Results follow focus. He made her transition a project in and of itself. Each day she sat with the engineers and data scientists, walking them through her workflow, one excruciating step at a time. She documented which additional sources she had to gather, all the PDFs and Excel files she had to read, and the judgement calls she made when the data was messy. The team compiled a library of past reports, annotated with notes about why one version was good and another fell short. The definition of "good" that lived in her head was now on paper and available to the whole team.

Document What Success Looks Like

To get an AI agent to work with you rather than against you, start by clearly defining what success looks like. Documenting successes (oftentimes, a bank of documents that illustrate your internal gold standard) will enable you to show concrete examples to the AI (often referred to as n-shot learning) and allow for context engineering to ground the AI in your specific business context.

At the same time, Ted had the engineers bring forward the flawed outputs from their first round of AI experiments. They laid the bad reports out side-by-side with the good ones and called out the failure modes: hallucinated numbers, misinterpreted code names, and boilerplate where nuance was needed. Naming these risks made them real and measurable.

Discuss Risk Collaboratively and Upfront

While very powerful, AI is not inherently safe. Discussing risk upfront is the only way to ensure the final solution is trusted. Documenting failure will allow you to judge the outputs of your experiments to ensure steady progress. Documenting what could go wrong will also help create AI guardrails that will protect against privacy leaks and security vulnerabilities. Take time to reflect on what's at stake for your organization. Which risks matter most to you?

- Brand damage from incorrect or misleading AI outputs
- Compliance, privacy, or security breaches
- Internal trust and change management failure
- Our data could be used to train proprietary models we do not control
- Infrastructure strain (latency, throughput, consistency, replication)
- Model behavior is too unpredictable or opaque
- Data quality gaps / insufficient data assets
- Lack of alignment with company voice or tone
- Ballooning / unpredictable costs
- Other: ____________________________________

Prepared to Succeed: A Disciplined Approach

By the time the analyst turned in her laptop on her last day, Ted had something they'd never had before:

- A clear statement of the problem they were trying to solve
- A shared understanding of what success looked like
- A clear understanding of where employees and AI excelled in the workflow
- A checklist of what must not go wrong.

It wasn't a solution in the form of an AI agent yet, but it was a clear path forward.

Process Re-architecting & Humans in The Loop

It's tempting to think AI agents can fully automate a process, but today's AI can't. Agents often fail beyond a couple of steps. When redesigning workflows, it's critical to map where humans must guide, supervise, and approve, and where AI truly adds value. Klarna's case study later in the book shows what happens when this balance is ignored.

Experiment, Iterate and Learn

Ted had seen enough false starts to know tinkering or vibe-coding wouldn't cut it. If his team was to make progress, **they needed to be more disciplined**. He introduced a new rule: no more vague experiments. Every test and trial needed to start with a clear hypothesis, a formal task definition, a set of metrics to measure success, and enough examples to contextualize the ask. He decided to **invest in an AI-to-business-impact platform to keep his team focused, accountable, and aligned**.

Constructing Good Experiments

- First, write down a **hypothesis**: what do you expect the AI agent to do, and why does it matter to the business?
- Next, define the **task**: is this *summarization*, *Q&A, translation, data transformation*, etc?
- Then identify **metrics**: how will you know if it worked?
- Finally, provide **examples** that will allow you to show the gold standard of what "good" looks like, and the pitfalls to watch for.

With those elements in place, the only remaining choices are which model to use and what prompt protocol to run.

Ted's first question was simple but central: *Could AI somehow accelerate the days-long work of reading through PDFs and Excel spreadsheets?* The team framed it as a **research** task followed by a **summarization** task. Their hypothesis was that the model could find and digest documents faster than any human business analyst.

The summarization half worked beautifully. Given a set of documents, and a small vector database containing the curated gold standard examples, the model surfaced all the key numbers and context the pipeline needed. What used to take days now took minutes. On the other hand, the so-called "research" was a mess. No matter how they tuned it, the model couldn't keep up with the pace of industry news. Ted took the lesson in stride: the AI could accelerate parts of the task, but it could not replace a human analyst's savvy and domain awareness.

The second round of experiments targeted another bottleneck. For every report, the business analysts pulled together raw data files. Because each report was heavily tailored to the customer, each set of files had a slightly different schema, and the data scientists had to handcraft SQL queries to fit the pipeline. The handoff was messy and it sometimes took days for a data scientist to be available to write the queries and kick off the next phase of the process.

The team framed this as a **query generation** task. Their hypothesis was that an AI agent could interpret the raw schemas and generate the necessary queries automatically. Using the model context protocol, the engineers built an integration for the agent to read schema definitions from the database and propose queries. To everyone's surprise, the agent was really good at generating queries that successfully mapped all the key fields correctly. Based on their experimental results, they estimated it could keep the pipeline flowing with no human intervention over 95% of the time, cutting out at least two more days from the process.

The team experimented briefly with having the model upsert entire database tables, but the platform team pushed back. They wanted to keep the schema-building in traditional code where they had full control. Ted agreed; automation didn't mean ceding all responsibility to some artificial brain. AI could handle routine query generation, while the platform engineers owned the infrastructure.

These experiments changed the mood of the project. All results, whether positive or negative, were now presented to the entire team. Ted was learning that **good experiments were about discipline and accountability**.

At the same time, the team was learning to find where **AI added speed and narrow down the places where human expertise was indispensable**. They were producing results they could trust and reproduce. It was time to put those results into production.

Klarna: A Cautionary Case Study

Klarna, a "Buy Now, Pay Later" (BNPL) financial technology company, made headlines in early 2024 when it replaced its customer service team with AI chatbots, claiming the chatbots did the work of 700 customer service agents. One year later, it **reversed course and re-hired employees** to replace the chatbots.

What happened? The AI chatbots failed to solve both routine and complicated support cases with empathy, which is needed when handling sensitive customer financial data. Customer churn spiked, so the bots had to go.

Lesson learned; **AI will fail without disciplined implementation**.

Translate Results to Actions to Outcomes

The summarization task was the first to make the leap to production.

The **summarization agent could take a stack of documents and spreadsheets and extract the key terms and figures in mere minutes**. The analysts were thrilled and began using it right away, copying the experimental prompt and pasting it into the enterprise AI chat tool whenever they were working on a new project.

Ted worried that if every analyst was allowed to keep their own version of the prompt on their work laptop, the process would fracture again. So he drew a line: there would be one prompt for this task, and it would be version-controlled and visible to everyone. The engineers wrapped the prompt and the chosen model into a deployed endpoint so no one had to paste in the prompt anymore, only call the endpoint, which would perform the summarization and send the results to the pipeline.

“Production” Means Realized Value

Getting real value from AI starts when AI agents move from prototypes into production. That’s the moment they deliver measurable results i.e. lower costs, greater productivity, and new capabilities. But “production” means different things to different organizations, so it’s important to define what it looks like for your business. The key question is simple: *what level of packaging and integration is needed for your AI agent to function as a real part of the team?* The right answer sets the standard that ensures your AI efforts move beyond experiments and into lasting business impact.

They used the same pattern to deploy the **query-construction agent, which could reliably turn raw data into valid SQL queries, saving days of data scientist time**. Within a few months, it was fully documented, monitored, and handed off to the infrastructure team that owned the final stage of the pipeline.

Use as Little AI as You Can Get Away With

Machine learning and AI are good tools for tasks that require flexibility, but in applications, most problems require the opposite: stability and predictability. **Resist the urge to just let AI do everything**; in applications this is often called “vibe coding.” But allowing an AI tool to generate the entire code base and underlying architecture, in addition to achieving the target task, is a recipe for building unreliable and unmaintainable code that will not survive production. Instead, **ensure your AI agents are scoped to a set of well-defined tasks.**

By the end of the quarter, their proofs-of-concept had become working systems. The analysts no longer spent days reading documents and instead focused on what they did best – keeping their ears to the ground and staying dialed into the customer’s business context.

By implementing both AI agents to augment his team, **a critical revOps process that previously took five people two weeks to make a single report now took one FTE less than 2 hours.**

Rather than handcrafting queries to maintain a brittle process, the data scientists could focus on building new experiments. Ted still had to remind them from time to time that each new experiment should have a clear business objective and a measurable outcome (and that no experiment should take more than a few weeks).

Meanwhile, the platform team was tracking API calls as well as token usage and error rates for the two deployed agents. As soon as anything went wrong, the engineers were alerted right away.

Pretty soon, Ted and his fellow execs were no longer getting called into Friday fire drills anymore, and he was setting his sights on a brand new problem-to-solve.

An Intuitive Approach to Cost Considerations

When estimating costs, Ted relied on a useful analogy: hiring full-time employees (FTE):

- **Role & Objective**s: Just as you write a job description, an AI agent needs a clear, business-aligned mission.
- **Onboarding**: Employees need training and tools; agents need prompts, context, and integrations.
- **Supervision**: Employees have managers; agents need monitoring, guardrails, and approval workflows.
- **Performance Reviews**: You measure employees against KPIs; agents should be tracked against ROI and business outcomes.
- **Cost Structure**: An employee has salary and overhead; an agent has platform, compute, and maintenance costs.
- **Scaling**: Hiring more people takes months; deploying more agents can happen in hours and work 24/7. Both require management.

Tips on Estimating Timeline and Cost

For cost structure, platform covers licenses and tools, compute is the cost of running models in production, and maintenance is the ongoing monitoring, updates, and support to keep agents reliable. A general rule-of-thumb:

- **Platform**: $1K–$10K per month (licenses, integrations, dashboards).
- **Compute**: $0.01–$0.10 per query, or $1K–$2K per month depending on volume and model size.
- **Maintenance**: 20–30% of initial build cost per year (monitoring, updates, retraining, support).

Defining Your Payback Period

Ted had to present to the executive leadership team and board, who mostly cared about one thing: the payback period. How quickly would the company recoup its investment in AI?

Ted calculated the cost of the status quo process: 5 employees took 2 weeks to complete the revOps process. He estimated their individual time allocated to the process over the two week period, multiplied it by their estimated hourly wage plus overhead, summed the costs, and made a reasoned, grounded monthly estimate of $5,000.

He then calculated the initial one-time build cost including experimentation and engineering hours to build the data pipelines at $20,000. He estimated inference usage, monitoring. maintenance costs, and platform costs at approximately $1,500/ month. He also factored in the 2 hours of analyst time it now took at $500.

Simple payback = One-time build cost ÷ Monthly net savings

Net savings = (status quo monthly cost) – (new monthly run-rate)

Applied to Ted's analysis:

- Status quo: $5,000/month
- New run-rate: $1,500 (platform/usage/monitoring) + $500 (analyst) = $2,000/month
- Monthly net savings = $5,000 – $2,000 = $3,000
- One-time build cost: $20,000

Payback = $20,000 ÷ $3,000 ≈ 6.67 months

Key Takeaways and Final Thoughts

Ted's story is not unique. All companies have smart people working in silos, maintaining processes that are profitable on paper but brittle in practice. Everyone is trying to figure out how to use AI to accelerate operations without degrading the services and customer relationships that underpin their bottom line.

Your organization has scope to gain new and valuable capabilities - AI agents are within your organization's grasp today. We said it in the beginning and we'll say it again:

Your "design space" to automate and innovate across your org is vastly larger than ever before, *if you have the vision*.

What Ted's story illustrates is that automation isn't about magic bullets. It's about acknowledging dysfunction (uncomfortable though that may be) and **doing the disciplined work** required to truly fix it and bring your vision to reality. That means:

- **Naming the problem** clearly and aligning it with business objectives..
- **Making a first effort** to solve it, even if it doesn't work.
- Using that first attempt to **clarify and scope** each use case.
- **Experimenting with discipline**: defining hypotheses, tasks, metrics, and examples.
- **Turning results into action** by making them usable, observable, and maintainable.

What's interesting to note is **how small a role AI models like OpenAI o4 or Google Gemini play in this story**, Ditto for Graphical Processing Units (GPUs). Models and GPUs are tools that serve a business need, like computers, and their value is derived **on how you adapt them to your business objectives.** See Appendix I: AI Use Cases by Domain for ideas.

New Revenue Opportunities Open Up

Not only did Ted see more cost efficiencies, he recognized **new value-added revenue streams that combined his organization's domain expertise with agentic capabilities** that were not possible before, specifically AI-powered managed services packages that extended existing support contracts. This got the sales team's attention who had been clamoring for new monthly recurring revenue (MRR) to add to current managed service agreements. He imagined the pitch: *"We already manage your IT environment. Now, we can safely embed AI into the same workflows to cut costs, speed up compliance, and reduce staff burden, all without adding new vendors or security risks."* For example:

- **Document Summarization-as-a-Service (DSaaS):** Automated summaries of data compliance reports, audit logs, vendor contracts, and clinical documentation.
- **Data Query Acceleration-as-a-Service (DQaaS):** "Self-service" business intelligence for analyst and compliance teams to quickly query data without technical expertise.

Horizontal vs Vertical AI

Commercial-off-the-shelf (COTS) solutions like Microsoft AI Co-Pilot can be valuable, but are not a differentiator. Not only can competitors use these tools, they offer diffuse, horizontal productivity gains (like employees writing marginally better emails) that won't really move the needle for your business. Moreover, COTS tools often have more to gain from you; their algorithms extract insights from your data to build models that become profit centers for them.

"Vertical," purpose-built AI agents that do real work are your path to a strategic differentiator to *deepen your moat* and realize new revenue streams. Focus on your domain expertise.

Now it's your turn. If you're ready to stop waiting around for magic and to start doing the real work of experimentation and learning, you don't have to do it alone. **Rotational Endeavor** was built for exactly this purpose. It's an AI platform that gives your team the structure, tools, and observability to run disciplined experiments, capture what works, and scale that into production.

The organizations that succeed with AI aren't the ones that buy the flashiest demos. They're the ones who **build the muscles that Ted built:** seeing dysfunction for what it is and holding himself and his team accountable for solving it. If you're ready to begin, **partner with us to turn one AI idea into measurable business value.**

Appendix I: AI Use Cases by Domain

Applying a functional lens can be helpful. The basic framework is to evaluate workflows in each function and determine the application of LLMs or GenAI tools to solve a real business problem. A common theme in these use cases is the ability for LLMs to unlock the value of unstructured data, often living in your knowledge base or data systems, at scale, opening new possibilities for visionary leaders.

Each use case should be evaluated by an impact metric, informing a cost-benefit analysis. An impact metric relates to improved productivity, cycle time, customer experience, brand, quality, and/or faster upskilling. The framework below shows how to scope AI use cases with discipline:

1. Name the problem
2. Define a hypothesis
3. Map to a well-scoped task
4. Define metrics
5. Determine context requirements

Marketing & Sales

Use Case: Brand Consistency Agent

In marketing, a recurring dysfunction is inconsistency: sales decks, collateral, and campaigns vary in tone and style, and every piece must pass through the marketing team before release. This creates a bottleneck that slows sales. One **hypothesis** is that an LLM prompt informed by brand guidelines could normalize tone and reduce approval time. The **scope of the task** would be style and tone evaluation: classifying whether drafts match the brand voice and revising them when they do not.

What must not go wrong is the model approving something that clearly violates brand voice or, worse, altering meaning while attempting to fix tone. Therefore, the **metrics** in this case are alignment checks: does the model preserve factual accuracy while adjusting tone? Does it avoid introducing new claims? Can it catch and correct "off-tone" drafts? To run such an experiment, a team would need **examples**, for instance, 10–20 pairs of collateral showing "before/after" brand-compliant revisions, along with 10 deliberately "bad tone" samples to test rejection.

The **business outcome** to measure is the average approval cycle and the incidence of brand violations. If collateral moves faster while staying on tone, the experiment has moved the organization forward.

Use Case: Product Feedback Summarization Agent

In many organizations, customer feedback is collected but sits unused until a quarterly review, delaying product improvements. One **hypothesis** is that an AI prompt informed by historical feedback could summarize open-text responses into themes and rank them faster than manual review. The **scope of the task** would be summarization and clustering: condensing responses into coherent themes and prioritizing them.

What must not go wrong is the model missing a critical pattern that customers are consistently raising. The **metrics** in this case are alignment checks: does the model correctly differentiate key themes or collapse problems into a "miscellaneous" bucket? Does it hallucinate new feature requests that aren't present in the data? Does it prioritize the right themes? To move forward with such an experiment, a team would need 20–30 **examples** of past survey datasets with human-coded themes and a curated list of product features.

The **business outcome** to watch is whether analysis turnaround time decreases without a loss of insight quality. If the company can act on customer feedback within weeks instead of quarters, the experiment has moved the organization forward.

Human Resources

Use Case: Voice of the Employee

HR teams often collect employee survey data but struggle to identify patterns and actionable insights quickly. One **hypothesis** is that a prompt informed by past survey data could cluster responses and surface important themes to the HR team. The **scope of the task** would be sentiment analysis and clustering: detecting tone and grouping related feedback.

What must not go wrong is the model dismissing negative signals or masking issues that would otherwise require urgent attention. The **metrics** in this case are alignment checks: does the model accurately identify sentiment without overgeneralizing? Does it surface themes that align with human-coded categories? Does it avoid trivializing serious concerns by grouping them incorrectly? The experimentation team would need **examples** of past surveys with human-coded sentiment labels and thematic clusters (bonus points for attrition data to correlate themes with outcomes).

The **business outcome** to watch is whether leadership can act faster on employee concerns without misrepresenting sentiment. If HR is able to respond within weeks instead of months and employees feel heard, the experiment has moved the organization forward.

Use Case: Onboarding Assistant

New employees often struggle to find the information they need in their first months, slowing productivity and lowering retention. One **hypothesis** is that a prompt augmented with company documentation, org charts, and policies could answer common onboarding questions consistently and reduce confusion. The **scope of the task** would be Q&A retrieval: surfacing authoritative answers from existing knowledge bases

What must not go wrong is the model giving new hires guidance that contradicts official policies. The **metrics** in this case are alignment checks: does the model provide accurate answers grounded in company-approved policies? Does it avoid fabricating procedures or referencing outdated information? Does it consistently cite the correct source of truth? To run such an experiment, a team would need **examples** of real onboarding FAQs paired with HR-approved answers, ideally with logs of past support tickets.

The **business outcome** to watch is whether new hires reach proficiency faster and require fewer manual interventions from HR. If onboarding feels smoother without creating compliance risk, the experiment has moved the organization forward.

Customer Service/ Success

Use Case: Customer Service Support Automation

Technical product companies often face a growing backlog of customer support tickets, many of which are routine inquiries. One **hypothesis** is that a prompt informed by product documentation and resolved tickets could resolve Tier 1 issues automatically. The **scope of the task** would be FAQ-style Q&A and classification: identifying the intent of a ticket and responding with appropriate information.

What must not go wrong is the model giving incorrect instructions that worsen a customer's problem or attempting to handle sensitive tickets outside its scope. The **metrics** in this case are alignment checks: does the model answer only within the scope of documented issues? Does it maintain brand tone and empathy? Does it correctly route non-routine issues to human support? To run such an experiment, a team would need **examples** of historical tickets with human-provided resolutions, plus gold-standard transcripts showing empathetic responses.

The **business outcome** to watch is whether backlog volume decreases while customer satisfaction remains steady or improves. If Tier 1 issues can be resolved automatically without harming experience, the experiment has moved the organization forward.

Use Case: Customer Sentiment Analysis Agent

Businesses sensitive to customer perception often lack real-time visibility into how sentiment is shifting across channels. One **hypothesis** is that a model could analyze customer feedback streams in real time and flag negative sentiment trends earlier than manual review. The **scope of the task** would be sentiment classification and trend detection.

What must not go wrong is the model failing to flag a genuine negative shift in sentiment until it's too late. The **metrics** in this case are alignment checks: does the model consistently identify negative sentiment without confusing sarcasm, mixed signals, or neutral comments? Does it avoid overreacting to isolated complaints? Does it surface meaningful trends instead of noise? For the **examples,** a team would need historic feedback with human-coded sentiment labels, plus churn data linked to feedback timestamps.

The **business outcome** to watch is whether churn or negative engagement decreases when interventions are informed by early warnings. If sentiment analysis enables faster, targeted response without overwhelming teams with false alarms, the experiment has moved the organization forward.

Appendix II: Rotational's AI Terminology

AI terminology is notoriously imprecise and subject to interpretation. It's important that we establish a shared terminology with our customers so we do not misunderstand each other.

Here are some key terms and how we use them:

- **Accuracy**: Used as a success metric in traditional supervised machine learning, but not always well-defined for LLM tasks. Accuracy works best when there's a single correct answer. For open-ended tasks, other metrics often matter more (e.g., helpfulness, tone, groundedness). Alignment, not accuracy, is often the north star in agentic development.
- **AI Agent**: A purpose-built operant powered by an LLM.
- **Context Engineering**: Structured process to gather the information the LLM needs to fulfill its goal.
- **Experiment**: A structured test to see whether an LLM-driven task performs well in your context.
- **Guardrails**: Engineered constraints and heuristics that reduce risk (e.g., safety filters, fallbacks).
- **Hypothesis**: A concise, testable belief about what an LLM could improve in your workflow.
- **Inference Server**: A hosted, persistent environment that hosts an LLM or computer vision model and generates AI responses when prompted. GPUs required.
- **LLM**: Large Language Models (LLMs) are generic models that can encode and simulate human language usage patterns.
- **Metrics**: What we care about measuring (e.g. tone, groundedness, cosine similarity).
- **Prompt**: A reusable prompt format to provide structured instructions to the LLM on how it should behave.
- **Qualitative Coding**: Using human judgment (ideally, internal experts at the customer's organization who are directly involved in the workflow) to tag inference server outputs. This allows us to move towards alignment.
- **Task**: A well-defined, granular task (e.g., question-answering, translation, summarization, classification) that an LLM has been fine-tuned to perform.
- **Test Cases/Examples**: Example inputs and expected outputs used to validate an LLM's performance.
- **Workflow**: A series of connected steps or actions your human team performs to get something done that has business value.

Our team has been implementing bespoke business automations using machine learning, data engineering, and AI for over a decade. Our logo is the lighthouse – a beacon in stormy waters to help travelers find their way to sturdy ground.

Endeavor is our enterprise AI agent development and deployment platform for creating trusted, business-aligned agents that deliver real results. It moves orgs from proofs-of-concept to tangible business assets through disciplined testing, measurable ROI tracking, strong governance, and seamless deployment. Orgs gain AI automation capabilities built on their domain expertise with full control and visibility, backed by security, compliance, and performance monitoring. It can be hosted or deployed on prem.