PREVIEW CHAPTERS

Emil Heidkamp

HORSEPOWER TO MINDPOWER

How AI turns cognition into an industrial process

Parrotbox.ai

Horsepower to Mindpower

How AI Turns Cognition into an Industrial Process

by Emil Heidkamp

Parrotbox.ai

Published by Parrotbox
www.parrotbox.ai

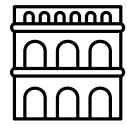
Cover and graphic design: Henry Casanova

Editing: Erick Piller and Weston P. Racterson, AI

Dedicated to

W. Edwards Deming and Margaret Hamilton,

who would both agree: we don't build them like they used to.



Concordes and Aqueducts

To be completely honest, some days I worry that my team and I are building horse-drawn carriages while someone else is inventing the airplane.

Our company, Parrotbox, builds AI systems that do unglamorous but important work for clients in a range of industries. We're not predicting how people will vote based on their grocery purchases or mapping the human genome to cure cancer. Instead, our AI agents coach bank employees to package loans or help social workers find jobs for people with disabilities. We spend weeks reformatting photocopied factory equipment manuals into structures AI can actually use. We write 55,000-word instruction sets so an AI financial advisor knows to ask the right questions before offering recommendations.

While we'll never win an innovation prize for solving the world's problems, these systems deliver results: achieving anywhere from 5x to 15x productivity gains versus unassisted human labor, in one case compressing a client's seven-month annual review of banking regulations into two weeks, with 10% higher accuracy than human reviewers…

And yet, despite these results, it still feels like we're behind the curve.

Every few weeks, one of the major AI companies releases something that leaves me shocked (if not surprised). A model that can reason through multi-step problems. An agent that can browse the web, write code, and execute it autonomously. A demo where someone says "Build me a complete project management application" and watches it materialize in real time.

Each new model is accompanied by big tech press releases suggesting that everything my team painstakingly builds for clients—the knowledge bases, the guardrails, the workflows and prompt engineering—will soon be unnecessary. Instead, organizations will simply ask the next model to "optimize our process for X" and the AI model will literally figure everything out on its own.

But then I think of the Concorde.

The Concorde was a supersonic passenger jet that could fly from London to New York in three and a half hours. It was a genuine marvel of engineering, and a glamorous status symbol for the billionaires and rock stars who flew on it. It was also commercially unviable, environmentally catastrophic, and was grounded permanently in 2003 after a horrifying, fiery crash that exposed numerous vulnerabilities in its design.

Meanwhile, all the boring subsonic planes kept flying, kept improving incrementally, and kept actually getting people where they needed to go.

The AI industry has a lot of Concordes right now. Bleeding edge AI systems that put on dazzling demonstrations of raw capability which make for extraordinary conference keynotes, yet many of these demos crash and burn under the pressures of real world enterprise implementation—where security requirements, regulatory constraints, and the simple expectation of reliable performance at scale render many exciting technologies commercially unviable.

So here we have a technology that I've personally seen do things that would have been science fiction five years ago, yet one that (according to MIT's State of AI in Business report) fails in large-scale, real-world deployments 95% of the time, a rate that would bankrupt an auto manufacturer or restaurant chain within a month.

How is that possible? What, exactly, are we supposed to believe about the current and future capabilities of AI and—more importantly—what are we supposed to do with it on an individual, organizational, and societal level today?

The answer is the subject of this book, and it has surprisingly little to do with what's happening at the current technological frontier.

---

[1] Aditya Challapally et al., The GenAI Divide: State of AI in Business 2025 (MIT NANDA, July 2025).

***

Our previous book was about AI for workforce training: using AI systems to create coaches, tutors, and support tools for employees. But a funny thing happened in the year after my team and I assembled that first book… Every client project that we had drifted beyond using AI for workforce training and toward building AI systems for augmenting, accelerating, or automating the work itself.

The conversations almost always followed the same pattern:

Month Zero:
"Can you build an AI agent to teach people how to do X?"

Month Two:
"Can the AI agent not just teach people to do X, but actually help them do X?"

Month Three:
"Do you think the AI agent could handle some or all of X on its own?"

It didn't seem to matter whether the client was in manufacturing, financial services, healthcare, government, or social services. Invariably, the client would have the same series of "A-ha!" moments and we'd go from AI agents preparing humans to do a job, to AI agents serving as copilots on the job, to AI systems doing the job with varying degrees of human oversight.

And while we welcomed the challenge, it certainly raised the stakes as we dove deep into the weeds of designing AI systems to do real work alongside real people, and we learned a lot. Some of it was genuinely new, some of it was timeless business / tech realities dressed up in AI clothing… almost none of it had anything to do with the hype-cycle AI discourse happening in the media.

***

To set some expectations, we won't be talking much about the capacities of specific AI models or how to connect OpenClaw to your work email, calendar, or DraftKings account. First, because those technical specifics will all change before this book is even released, and second because it's usually not the model or the specific agentic platform / harness that determines an AI system's real world success or failure.

Rather, we will discuss everything around the models and agentic platforms: how organizations structure their knowledge, how they design workflows, how they set performance expectations, how they monitor and govern AI systems, and—most fundamentally—how well they understand what AI actually is and what it requires to function reliably.

We remember the Roman Empire for the might of its armies, the architecture and spectacle of the Colosseum, and the Latin language that we use in science, medicine, and law to this day. But these achievements were all the fruits of Roman civilization, not the foundation.

Most people forget all the unglamorous but important things that made Roman civilization possible in the first place: starting with the aqueducts, pipes, and sewers that made a city of a million people possible when every other settlement on Earth topped out at tens of thousands. And the same could be said for how Romans mixed the concrete for their buildings or their elaborate system of quality control for the bread sold in markets.

These everyday innovations weren't a footnote to Roman civilization. They were Roman civilization. Everything else was built on top of it.

This book is about AI's plumbing, and all that it makes possible.

The ABC's of AI

A friend of mine once said "By the time you get done defining terms… nobody cares."

In that spirit, I'll keep this glossary section brief.

The language of AI is evolving so fast that the definitions below don't exactly match the definitions from our previous book. That said, here are the definitions we're going with this time around:

- **AI:** AI technology in general, though we'll mainly be talking about "generative AI" (the kind used by ChatGPT, Claude, Gemini to generate recipe recommendations and analyze spreadsheets) as opposed to the other types of AI used by self-driving cars and currency trading platforms.

- **AI Model:** The part of an AI system that analyzes input and generates an output—what most people would casually call "the algorithm." It's the component that makes ChatGPT behave differently from Claude.

- **AI Agent:** An AI-powered application designed to do specific things, with varying degrees of autonomy. When we talk about "AI coaches" or "AI preventive maintenance copilot" those would be examples of specialized AI agents.

- **AI System:** Any system that makes use of AI to accomplish work. While it might consist of a single AI agent, it could encompass many agents interacting with shared databases, tools, user accounts, security frameworks, and integrations with other systems.

All that said, sometimes I might lazily say "the AI" in reference to any or all of the above—and for that, I apologize.

Chapter 1

One Weird Trick: How AI Does What It Does

"Chaos is a name for any order that produces confusion in our minds."
— George Santayana

***

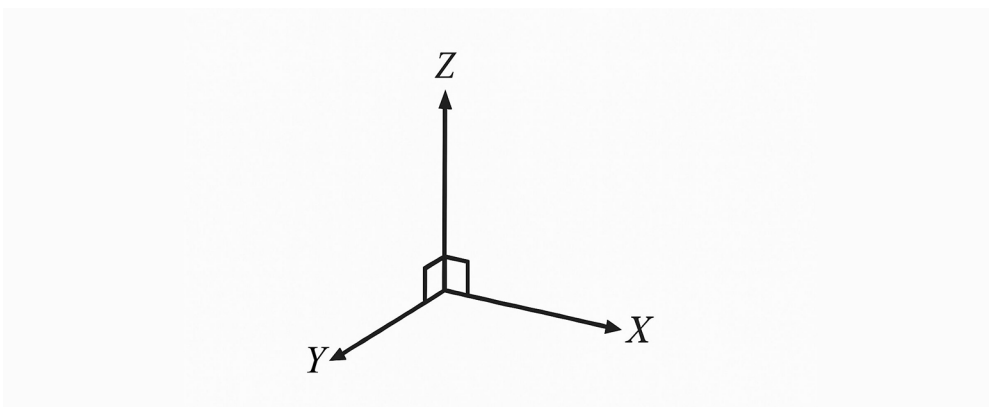
My mother was a math teacher. If there was ever a day when my sister or I didn't have any math homework, she would invent some. It wasn't fun on a summer day when you could hear your friends outside playing, but at least now I know my sines from cosines and logarithms from algorithms.

One day I asked my mother "What's the hardest math you ever learned?" She replied "In one of my university courses we did higher dimensional math. For instance, in higher dimensions, you could turn a basketball inside out without breaking it."

"How does that work?" I asked.

"Think of a three dimensional graph with X, Y and Z axes…"

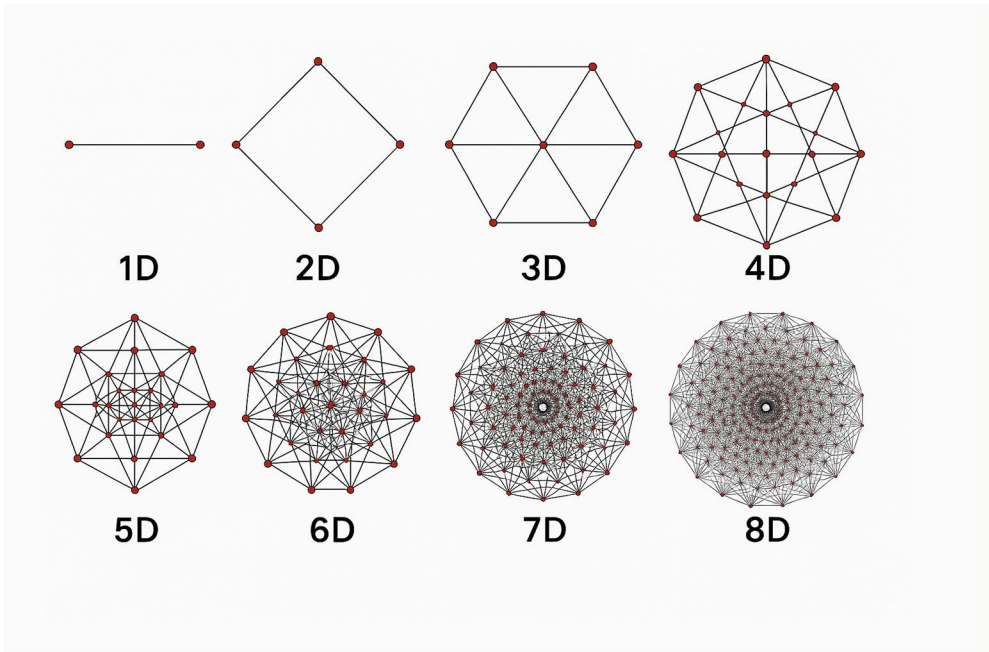
I did.



“Now imagine adding a fourth dimension that is at a right angle to each of the other three.”

I couldn’t.

“While that might seem impossible in the physical world, or in our minds, it’s entirely possible with math. In fact, you can keep adding new dimensions pretty much forever.”



“That’s amazing!” I said.

“It’s interesting…” she agreed. “But not terribly practical.”

Welcome to the Nth Dimension

It turns out civilization now runs on the exact kind of math my mother dismissed as “interesting, but not terribly practical.”

Molecular biologists use higher-dimensional math to fold proteins for Alzheimer’s medications. Telecom engineers use it to keep cellular signals clear. And, of course, it forms the foundation of modern generative AI.

This became possible because we now carry the equivalent of 1980s supercomputers in our pockets. There are tens of thousands of computers flying above our heads inside Starlink satellites. And when it comes to AI, we are building computing centers in the desert the size of Manhattan.

All this so you can pull out your phone in the middle of the wilderness, ask, "How can I tell a mosquito bite from a tick bite?" and receive an answer from an AI model running on the other side of the planet.

Underneath that seemingly simple interaction are computers navigating mathematical spaces with thousands of dimensions: spaces humans still cannot truly visualize, despite how commonplace the technology has become.

How This Powers Your Chatbot

Because AI models process natural language input and produce natural language output, it's tempting to assume they relate to language and information the same way humans do. However, in reality AI models don't "speak," "read," or "understand" language: they map it.

Imagine every word, phrase, concept, and association in human language existing as points in an impossibly large mathematical landscape. In this space, words that are related sit closer together, but in countless different dimensions. Which word is "closest" depends on which dimension you're looking at.

For instance, if we started from the word "strawberry" we could go in various directions such as:

Strawberry… ice cream
Strawberry… blonde
Strawberry… exports from Mexico are down this year
Strawberry… Fields Forever

We still don't understand how humans can tell that "strawberry" means something different in each of these sentences. However, we've managed to approximate that intuitive understanding mathematically with modern AI models.

When deciding what word should come next, the model analyzes the entire conversation up to that point, mapping the relationships between words, phrases, suffixes, prefixes, and punctuation marks ("tokens") like a higher dimensional connect-the-dots.

The model then compares those patterns to similar patterns it learned while processing trillions of words of "training data," back when the model was initially developed. Based on this, it will decide whether the current conversation containing the word "strawberry" more closely resembles discussions about desserts, hair color, agriculture, or Beatles lyrics.

Based on those similarities, the model predicts which next token is most likely to make sense in context.

The AI model performs this process not just for the word "strawberry," but for essentially every token in the conversation, every time you send a message. And that's why we need data centers the size of Manhattan to process all of our AI interactions.

From Weird Tricks to Useful Work

I'm not a fan of Ayn Rand, the philosopher. But while her ideas about capitalism and how to organize society are kind of scary, her definition of knowledge work is useful when discussing AI:

> ***"[Any] resource which, in order to become of use or value [to humans], requires the application of human intelligence, is not a 'natural resource'… but a product of the human mind."***

Building on this, we might say:

> ***"AI creates value by taking [largely] useless information and making it more useful [to humans] through the application of machine intelligence."***

Just as you cannot put crude oil directly into an automobile without processing, a stack of 10,000 customer service emails is not inherently valuable. Neither is a 400-page banking regulation manual, a decade of factory maintenance logs, or a folder full of medical intake forms. In their raw state, they are mostly noise: too large, repetitive, and fragmented to offer any actionable insight without processing.

Continuing the information-as-oil metaphor, AI systems create value by acting as refineries, identifying patterns inside that noise and transforming it into actionable knowledge products: summaries, recommendations, predictions, classifications, decisions, or drafts.

In other words, while we might like to think of AI systems as magical digital employees (because they speak our language), in reality they are more like digital industrial infrastructure.

And just like an oil refinery, the quality of what comes out depends heavily on the design and precision of the machinery, and the quality of inputs that feed it.

A Tour of the "AI Factory"

Recently, my company built an AI agent to help businesses identify opportunities for energy efficiency improvements. The agent was taught to analyze businesses in terms of their inputs (e.g., cocoa beans delivered to a cocoa plant), processes (workers feeding the beans into roasters, grinders, and mixing tanks), and outputs (powdered cocoa, ready for use in cakes and chocolate bars).

That particular AI agent was designed for non-technical stakeholders (specifically bankers), providing just enough information at just the right level to determine whether a factory boiler upgrade or swapping out diesel vans with electrics would qualify for World Bank climate funding.

That's a good metaphor for how much business stakeholders need to know about AI systems to implement them successfully. You don't need to understand transformer architectures, gradient descent, or tensor mathematics to successfully implement AI systems. But you do need to understand their basic inputs, processes, and outputs, at a high level.

Inputs

For generative AI applications, the inputs would include:

Context (user-supplied information): This includes the transcript of the current conversation (user input, AI output) plus any additional information being fed to the AI (e.g., from database queries, file attachments, integrations with other systems).

Core Training (the AI model's pre-existing memory): This refers to the data originally used to train the AI model, allowing it to map relationships between words, concepts, and patterns that inform future outputs.

While both types of input are essential, just like in a cocoa factory they introduce a risk of cross-contamination. For example, while developing a ChatGPT-based AI agent to help social workers find jobs for people with disabilities, we ran into situations where the agent would occasionally include six-month-old expired posts that weren't even in the sets of job listings we'd been feeding it.

At first this seemed like a random "hallucination," but then—on closer inspection—we found that the AI model was including old jobs with the same titles at the same companies as the jobs in our data set. The AI model wasn't including those listings randomly, it just literally couldn't distinguish between the data it had been trained on before and data we were handing it now.

While, eventually, we were able to implement measures to prevent old listings from appearing, there's always a risk of that kind of thing in AI systems. Basically, it's the AI equivalent of a candy wrapper allergy warning stating a chocolate bar was "Made in a facility that also produces tree nuts." Whether your organization can safely consume it depends on how allergic you are to errors.

Processes

Once an AI model receives input, it will perform various operations, including but not limited to:

Inference: This is the technical term for the process we discussed earlier, where an AI model compares the patterns in the user's input to all the patterns in its pre-existing memory then predicts which "token" comes next (e.g., "strawberry ice cream" vs. "strawberry blonde").

How reliably, quickly, and cost-effectively an AI model can predict the next token is the primary performance metric that big AI companies like OpenAI, Anthropic, and Google compete on.

Orchestration: As mentioned above, different AI models have different performance specifications. ChatGPT 5 is generally "smarter" than GPT

5 mini or GPT 5 nano, but it is also slower and more expensive. Likewise, not all AI tasks are equal: some tasks like "Write a <150 word bullet point summary of the issue" may require far fewer instructions compared to "Guide the safety inspector through the 30-point evaluation framework."

Complex AI systems need the ability to switch between models and load different instruction sets ("prompts") depending on the task, like an automobile switching gears.

Retrieval: This refers to an AI system's ability to proactively import additional context needed to give a response or make a judgment. This could include a chatbot doing a web search before answering a question about the stock market or pulling in the complete documentation for a factory machine before telling which type of oil to pour into the engine.

There are many ways an AI system can do this, each with their own pros and cons, which we will discuss in later chapters.

Context Management: Modern AI models can accept a staggering volume of input, in some cases millions of tokens / words. However, their ability to pay attention to that input is finite. Giving an AI model a 28,000 word corporate policy manual and expecting it to give an accurate answer about whether employees may keep parakeets in their cubicles in less than 300 milliseconds might be asking too much.

Thus, effective AI systems need a mechanism to selectively un-load or compress contextual information when it is no longer needed, to focus attention and improve speed, quality, and / or cost.

Integrations and Tool Use: AI systems do not exist in isolation. To complete tasks, they often need to transmit information to and from other systems (like databases, a sales team's CRM, or a factory's maintenance management system) or issue commands to traditional software applications ("Create a new record in the spreadsheet," "Activate the camera.")

How difficult it is to set up these connections and how reliably they operate depends on the specific systems involved, the quality of the instructions and input, and the innate capabilities of the AI model.

Outputs

Per our previous metaphor of AI as an information refinery, the outputs of an AI system could include:

Knowledge Products: A report on a project team's progress versus spending, a set of lightly personalized marketing emails, an assessment of the severity of a natural disaster based on updates from the field, catering suggestions for an office holiday party.

Decisions and Actions: Category tags correctly applied to records in a product database, potential fraudulent activity reported to a bank's financial and economic crimes team, a lunch reservation at your favorite restaurant. How to measure, monitor, and control the quality of these outputs will be a major concern later in this book- and likely a major concern for human civilization in the decades to come.

The main challenge is that, unlike traditional software quality assurance where an app either works or it doesn't, [2] an AI system is more like a factory where employees make mistakes and machines drift out of calibration. In this world, the goal isn't to achieve perfection, but to keep the system within tolerances through constant vigilance, effective guardrails, and continuous improvement.

AI Is Not Software (As We've Known It)

If you take away anything from this overview of generative AI technology, appreciate that AI models and systems don't behave like traditional software. Where traditional software executes specific instructions, AI navigates messy webs of relationships between words and concepts. Where traditional software is driven by clear yes / no logic, AI operates in percentages and probability.

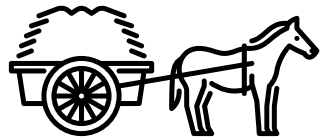
While that might seem strange and new for computing, it's not so strange for industry. We already have time-tested models for taking chaotic, error-prone, probability-driven entities and somehow getting them to do useful work. We call it "managing people."

But what does it mean to manage an organization where not all "people" are "human"?

---

[2] Setting aside issues of scale and infrastructure like data pagination or server load balancing.

Chapter 2



Bots of Burden: From Artificial Intelligence to Real Work

"[Show] me a creature that thinks as well as a man or better than a man, but not like a man." —John W. Campbell Jr.

***

My friend Quince used to lead a double life writing for the New York Times and running a horse farm in Wisconsin.

Once, back when I was a freelance journalist capable of working from anywhere, Quince called me in the dead of winter and said, "Look, I need help transporting horses, and you need to learn how to handle animals and drive stick shift. So, are you coming up here or what?"

I didn't accomplish either of those goals, but that's okay: I think Quince just wanted some company on the road while hauling his horse trailer.

When we got back to the farm, it was time for Quince's annual holiday party, which attracted an odd mix of his New York literary and art world friends, his rural Wisconsin neighbors, and his old army buddies. At one point a group of us ended up in a hot tub sipping tropical drinks during a light snowfall.

In the distance, you could see Quince's shepherd dogs watching over the sheep that he raised for wool. While hanging out on the farm, it was so common to pass a sheep or a shepherd dog on the dirt trails that you just stepped to one side and let the animal pass without thinking about it.

But on that particular day, a sheep and one of the dogs crossed paths directly in front of the jacuzzi. And right as we all were talking and laughing, the dog suddenly lunged for the sheep's throat. Blood splattered across the snow like something out of an especially violent crime movie.

"What the hell just happened?!" I said, bolting out of the hot tub into the freezing air.

"I don't know," Quince said, already going back inside to get his work clothes. "But we're having mutton… and selling a dog."

To this day I will never know what flashed through that dog's head when, after about seven years of peaceful coexistence, she finally decided to take a bite out of one of the sheep. According to Quince, it's not entirely unexpected on sheep farms, but once it happens the dog will never be fit to work again.

Some days, AI systems remind me of that dog.

Harnessing AI

For thousands of years humans have raised, trained, and used animals for work to take advantage of their superior or complementary abilities. Dogs can track prey animals from a distance, watch over flocks, and detect and fight off predators or rival human groups. Horses can pull heavy loads and carry human riders over long distances at high speed.

Neither dogs nor horses are 100% consistent or predictable, and neither is capable of explaining their actions, decisions, and occasional freak-outs after the fact. But they can do useful things humans can't.

The same is true for AI. When it comes to capabilities:

- **AI can process vastly more information than humans.** A human reads one report at a time, at about 200 words per minute. An AI system can analyze thousands of documents, emails, maintenance logs, or regulations simultaneously, in minutes if not seconds.

  For one financial sector client, our team created an AI agent that condensed their usual seven-month regulatory review process into two weeks, with higher consistency than human reviewers. Spot checks found the AI's output accurate more than 94% of the time on the first pass, versus roughly 74% to 82% for human reviewers.

- **AI can perform repetitive cognitive work continuously, at high speed.** Humans get bored, distracted, and fatigued. AI systems can repeat tasks involving basic "understanding" and straightforward "judgment" indefinitely with near-identical output quality, without the need for sleep, weekends, coffee, or recovery time after back-to-back meetings. This could include adapting marketing copy for different segments, categorizing records, or reviewing recorded calls for potential compliance issues.

- **AI can recall and apply knowledge instantly.** Humans forget details, overlook policies, and struggle to retain large bodies of information. One of our manufacturing clients told us, "Everything in our factory is documented, but it's documented in 500-page manuals nobody reads." We were able to build an AI support agent that allowed technicians to ask natural language questions like "What oil does this compressor require?" or "What's the lockout procedure before servicing this machine?" and receive answers in seconds, drawn directly from the company's existing documentation with citations.

- **AI is available to everyone, 24/7.** Human expertise is often bottlenecked around small groups of specialists who simply cannot be available to everyone in their organization on demand.

  To give an example, our company created a coaching program for financial advisors that originally consisted of two small-group coaching sessions per month plus online discussion boards monitored periodically by human instructors. In the latest version of the program, we added an AI "virtual coach" capable of answering questions, roleplaying important client meetings, drafting outreach emails, and helping tailor seminar decks for specific audiences, at any hour of the day.

In short, AI allows many forms of "knowledge work" to operate around the clock at machine tempo rather than human tempo, much like factories, logistics networks, and electrical grids already do today. And the long term effect of this will likely be customers expecting next-day delivery from law firms and engineering departments the way they already do for Amazon deliveries.

And even those advantages still operate on a "tactical," task / process level. AI has less obvious but arguably more profound capabilities that emerge when applied strategically, at scale. For instance:

- **AI can dramatically improve visibility into operations,** monitoring and summarizing activity on an individual, team, department and organizational level without anyone having to stare at a dashboard, wondering what all the graphs and numbers mean. It can also pick out exceptions that might otherwise escape notice. Our company is already doing this for clients. In one case, we built an AI system used by 150 bankers and the system was able to give plain-language reports on the entire group's performance while also noting, "A loan officer in Guatemala seems to be gaining traction with agricultural clients… perhaps someone from the regional or global office could reach out and support them."

- **AI can coordinate work and allocate resources across complex, rapidly changing systems.** [3] Organizations have been using computers to monitor output, forecast demand, and track inventory since the 1960s. AI takes this to the next level by allowing computers to identify patterns and account for factors that cannot be easily translated into conventional math, such as hospitals estimating likelihood that a patient will no-show for an appointment, or [4] Wal-Mart adjusting sales forecasts for products based on social media trends. [5]

- **AI can simulate decisions before humans implement them.** Modeling the consequence of decisions is one of the primary advantages of human intelligence ("What will happen if I try to sneak past the tiger?") but usually breaks down at institutional scale.

  Back when I worked at a healthcare financial consulting firm, we helped hospitals model their response to possible future scenarios (for example, how a shift toward maternity care versus a cancer center would affect the bottom line). Setting these models up took effort, but the work was relatively straightforward because every input was quantitative.

---

[3] MaryLou Costa, "Would Having an AI Boss Be Better Than Your Current Human One?," BBC News, July 3, 2024, https://www.bbc.com/news/articles/c03lgz2zrg1o.

[4] Yousif Mohamed AlSerkal et al., "Real-Time Analytics and AI for Managing No-Show Appointments in Primary Health Care in the United Arab Emirates: Before-and-After Study," JMIR Formative Research 9 (2025): e64936, https://doi.org/10.2196/64936.

[5] Tensor Strategy Group, Case Study: AI-Enabled Demand Forecasting and Inventory Optimization in Retail, https://www.scribd.com/document/925621236/Casestudy-Walmart.

AI now lets us model far messier situations. For example, our team worked with the cybersecurity staff of a streaming video platform to build a multiplayer, AI-facilitated simulation in which everyone from the CEO to the legal department to the customer service team could rehearse their responses to a major data security breach, practicing the exact wording they would use in conversations with employees, government regulators, and concerned customers.

- **AI can turn organizations into continuous learning systems.** In our work advising organizations on workforce training and knowledge management, we've found that "post-mortem" reviews (now rebranded as "learning sessions") are the one item on a project plan most likely to be skipped.

  By the time a project formally closes, people are exhausted and have usually moved on to other work after their individual contribution wrapped. And while someone might drop a comment into a report saying "This was a great innovation: we should make it part of our SOPs," those insights typically languish in a SharePoint or Google Drive folder, never seen by human eyes again.

  However, now that AI gives us a new set of eyes, we might finally be able to retrieve and learn from all those scattered flashes of insight and innovation at the grassroots level.

Individually, none of these capabilities are entirely new. Organizations have long used databases, dashboards, workflow automation tools, predictive analytics, and search engines to monitor work and augment decision-making. But when you layer on AI's ability to spot patterns that a deterministic dashboard wasn't originally set up to track, derive insights, and even decide how the organization should act on those insights.. the effect is transformative.

A law firm where AI systems review contracts and draft summaries overnight, monitor regulatory changes continuously, and look up relevant cases instantly is not simply a "faster" version of a traditional law firm: it's a fundamentally different type of organization delivering a fundamentally different service.

Measuring "Mindpower"

To appreciate the significance of AI as a new source of cognitive labor, it's useful to compare it to the only other source of cognitive labor we know: the human brain. And, handily, we already have a conceptual model for comparing the productivity of organic beings to machines: horsepower.

When James Watt needed to sell steam engines to mine operators in the 1780s, he didn't talk about pressure differentials or thermal efficiency. He talked about horses. Specifically, how many horses his machine could replace. A draft horse generates about 1 HP of sustained output. A Ford F-150 generates 400. A Formula 1 engine pushes 1,000.

Today, nobody expects a horse to outrun or out-haul a motor vehicle on the highway. And, if we're going to have an honest conversation about the implications of AI for human jobs and society as a whole, then it's time we applied the same brutal honesty to cognitive work.

Let's call our new metric mindpower. If we peg median human cognitive output at 1.0 MP (the baseline for an average knowledge worker doing average knowledge work) then where do most humans land?

No matter how you calculate it (standardized test scores, reading speed, written output) the math is humbling. Even if we used IQ as a lazy proxy for human cognition and knowledge work productivity, if the median IQ is 100 and only 2.2% of humans exceed 135, then a profoundly gifted or genius-level person with an IQ between 150 and 200 (the top one tenth of one percent of human intelligence) would still only rate 1.5 to 2.0 MP.

That makes the world's smartest people the cognitive equivalent of:

- A motorized pump in a residential above-ground pool
- A portable workshop air compressor
- A small table saw
- A go-kart engine
- A trolling motor for a fishing boat

The uncomfortable truth?

Even in our own company's modestly scaled projects, we've already clocked AI-augmented human workers and autonomous AI systems operating between 5x and 10x human productivity and beyond for highly repetitive tasks.

If present trends continue (a dangerous phrase, but warranted here), no human lawyer in 2035 is going to give you a better explanation of international trade regulations than an AI agent running a 250 mindpower Gemini 21 model optimized for legal research. That would be like Secretariat trying to outrun a Lamborghini, or a pack mule pitted against a pickup truck.

Today's organizations evolved to cope with a scarcity of cognitive capacity, where humans get tired, forget information, require years of training, communicate slowly, and can only pay attention to a limited number of things at once. Entire management hierarchies, reporting structures, workflows, and professional specializations evolved around those constraints.

But if AI can turn cognition from a constraint into an abundant commodity, then the evolutionary constraints that shaped our institutions disappear, and new species of organizations will evolve to take advantage of the new climate.

When the Dog Bites

Most of the time, working with AI feels surprisingly… normal.
You type a question in plain English. The AI responds in plain English. It understands context, picks up on nuance, adjusts its tone to match yours. If you've ever had a good conversation with a knowledgeable colleague who just gets it, that's what interacting with a well-designed AI agent feels like. It's easy to forget you're not talking to a person.

And that's the problem.

Because AI, like Quince's farm dog, is still an alien intelligence. It responds to verbal commands, but it doesn't "think" the way humans do at all. And the more comfortable you get with it, the more shocking it is when something unexpected happens to break that comfortable rapport.

For example, once when we were reviewing outputs of a system that helps bankers write loans for renewable energy projects, we came across a case where a banker in Côte d'Ivoire had a technical question about solar panels, and our AI agent suggested they contact the local chapter of the Syndicat des Énergies Renouvelables (SER), an international solar industry trade association based in France with offices across multiple Francophone countries.

Reasonable suggestion. Except our research found that SER didn't have a chapter in Côte d'Ivoire.

So was the AI agent lying to its user? Not exactly.

The AI model didn't invent a SER chapter in Côte d'Ivoire out of nowhere. It identified a plausible pattern connecting "solar power," "cost estimation," "French language," and "Africa," then confidently extended that pattern one step too far, assuming that SER must have a branch in Côte d'Ivoire like it does in nearby countries like Senegal.

In this case, the advice was still directionally correct ("contact your local solar industry association") and, it was easy enough to Google the correct organization (the Association des Professionnels des Énergies Renouvelables de Côte d'Ivoire). But the incident raised an uncomfortable question: How do you know when the AI is right and when it's confidently wrong?

This wasn't a bug in the traditional sense. The AI didn't malfunction. It did exactly what it was designed to do: calculate probabilities across billions of parameters and select the most likely next token. From the model's perspective, "there is an SER office in Abidjan" and "there is an SER office in Dakar" were both high-probability completions given the input context. The model doesn't know which one is true: it just knows which patterns of tokens tend to appear together in its training data.

This is the same thing that happens when you confidently tell someone "turn left at the gas station" and then realize, five minutes later, that you meant "turn right"—because your mental map of the neighborhood has those two landmarks reversed. You weren't hallucinating. You weren't lying. You just predicted the wrong word based on an imperfect internal model of reality.

The difference is, when a human does this, we recognize it immediately. We know what it feels like to be uncertain, to second-guess ourselves, to realize we might be wrong. AI models don't have that. They output every token with the same level of confidence, whether they're 99.9% sure or 51% sure. You can't tell, from the output alone, whether the AI is giving you verified fact or plausible fiction.

Quince's dog wasn't insane. She wasn't malfunctioning. She was doing what made sense to a dog, for reasons no human could fully reconstruct. Maybe the sheep moved in a way that triggered a prey drive. Maybe there was a scent in the air that signaled danger. Maybe it was something else entirely. We'll never know, because dogs don't think in words, and they can't explain themselves.

AI systems are the same way. And here's the uncomfortable truth: humans do this too. How many times have you heard someone say "I don't know what I was thinking" after making a catastrophic decision? How many times have you said it?

The difference is, we've had thousands of years to develop instincts for working with other humans. We know how to spot when someone's about to do something reckless. We know how to build systems (laws, norms, organizational structures) that account for human unpredictability. We've learned to trust people enough to work with them productively, without trusting them so much that we're blindsided when they do something inexplicable.

We don't have those instincts for AI yet.

Digital Domestication

We don't know how many Neanderthals got mauled trying to make friends with wolves or how many ancient nomads got trampled trying to climb on top of a horse. We do know that someone eventually figured it out, because we've been riding horses and having dogs herd our sheep for millenia. And we also know that it was a slow, iterative process.

In the 1950s, Soviet geneticist Dmitri Belyaev took wild silver foxes and selectively bred them for tameness. Within six generations, he had foxes that wagged their tails and sought out human contact. Within ten, he had foxes that were behaviorally indistinguishable from dogs.

But the first generation still bit. The second generation still bit. It took time and selection pressure and a lot of trial and error to get from "wild animal that might try to kill you" to "companion animal that reliably won't."

AI is our new dog, horse, and ox. It's powerful, it's useful, and it operates on logic we don't fully understand. Our job isn't to make it human (that's impossible); it's to learn how to work productively alongside it, and build systems that account for its strengths and its weirdness, and to develop the instincts that let us trust it enough without trusting it too much.

One might even say that AI system design is the new animal husbandry. We're learning to shape something alien into something that reliably serves human purposes. And hopefully we won't have to get bitten too many times in the process.

Chapter 3

Freeform, Transactional, Systemic: Three Levels of Human-AI Coordination

"It is the long history of humankind (and animal kind, too) that those who learned to collaborate and improvise most effectively have prevailed."
—Leon C. Megginson

***

For most of human history, coordination was simple.

If a hunter-gatherer band spotted a mammoth, coordination amounted to "You chase it into the canyon, I'll jump on its back." No workflows. No steering committees. Just a few humans and their dogs, those same domesticated wolves from the last chapter, improvising in real time to bring down whatever animal happened to be in front of them.

As societies grew more complex, that stopped working. People specialized. One person knapped flint all day, developed a consistent process, and taught it to their children. They traded with the person who had figured out barley. Coordination became more focused, more repeatable, more transactional.

Then cities emerged, and empires after them. Projects like the pyramids required coordinating thousands of laborers and artisans who would never meet one another. Entirely new machinery became necessary: managers, administrators, written processes, reporting structures, systems of governance capable of aligning enormous numbers of people around a single mission no individual could see the whole of.

We just spent two chapters establishing what AI is—an alien intelligence you harness rather than instruct, a refinery that turns raw information into useful product. This chapter is about the next problem, the one almost every organization gets wrong: now that you have this new kind of worker, how does it coordinate with the humans you already have?

Because if you look at AI in the long sweep of management history, we are doing something faintly absurd, onboarding an entirely new workforce into organizational structures that took ten thousand years to evolve, and we're doing it in about eighteen months.

The results have been, predictably, a bit chaotic. At the top, executive committees debate which models to license and how to word their governance policies without quite knowing the shape of the monument they're trying to build. At the bottom, individual employees are handed a chatbot and told to "experiment," wandering the landscape like those prehistoric hunters with their dogs. Then everyone wonders why the productivity gains never quite arrive.

After building AI systems across financial services, social services, manufacturing, healthcare, and government, my team has come to believe organizations need to master three fundamentally different modes of human-AI coordination:

- **Freeform coordination,** where an individual human sets the agenda and the AI serves that agenda.
- **Transactional coordination,** where humans use AI to accomplish a predetermined task in a predefined way, and the AI serves the process.
- **Systemic coordination,** where humans and AI systems each hold roles, working together or independently, in service of the larger organization.

One caveat before we go further, because it's the most common way people misread this framework: these are not stages of maturity.

Organizations do not graduate from Freeform to Transactional to Systemic, any more than civilizations abandoned hunting once they learned to farm, or abandoned farming once they learned to build pyramids. Each mode solves a different coordination problem, and a healthy organization runs all three at once.

Freeform Coordination: Your Wish Is AI's Command

Most of our everyday encounters with AI are freeform.

When you open Claude or Copilot and ask it to draft a proposal, or to suggest a way to make pancakes without eggs, you are setting the objective, driving the workflow, and deciding what success looks like. The AI serves at your pleasure, within whatever broad guardrails the organization has set. It doesn't know or care about your company's standard operating procedures. It's just helping you solve the problem in front of you, however you prefer to solve it

.

Some amount of freeform coordination will always be necessary, because novel problems don't arrive with a pre-built workflow attached. When a logistics coordinator has to explain to a client why a shipment is stuck behind a port strike without sounding like they're making excuses, or a copywriter has to reconcile a newly acquired company's product catalog with the parent company's, the entire value of an AI collaborator is its flexibility. And sometimes a team member stumbles, through pure experimentation, onto something that should become standard practice.

In freeform mode, the questions that dominate mainstream AI discourse are actually the right questions:

- "Which model is the better technical writer, Gemini or Claude?"
- "How do I phrase this prompt so the AI flags risks I haven't thought of?"
- "Is it worth paying for the more powerful model to rewrite this query, or will the cheaper one do?"

These are real, live considerations when an individual is steering.

But here is the trap: organizations that operate only in freeform mode, never moving past prompting tips and tricks, will not see meaningful ROI at scale.

When an organization's entire AI strategy is "give everyone a chatbot and see what happens," what happens is what always happens with a new tool: a handful of enthusiasts become power users, most people poke at it occasionally, and the gains never compound.

Worse, when a power user does hit on something valuable, they're more likely to post it on LinkedIn than tell the colleague at the next desk. Meanwhile leadership waits for productivity to bubble up from the collective unconscious while costs quietly rise, especially among the heavy users in engineering and marketing.

In conversations with leaders everywhere from small businesses to Fortune 500s to the Big Four, I'd estimate that in well over 80% of cases the story is identical: "We bought everyone licenses and told them to experiment. If we want to use any other tool, we have to get approval from IT or the innovation committee."

I try to be polite when I hear this, but I've come to think of the pure "fool around and find out" approach as a peculiar cocktail of abdication and obstruction—the organization shirking its responsibility to design good systems while simultaneously blocking the grassroots experimentation that might have rescued it. It manages to be both too permissive and too controlling at once.

And it's worth noting that this isn't an AI problem. It's the same pattern that played out with the PC, the web, and the cloud: deploy broadly, let a few enthusiasts find the value, watch most of the capability go unused, and only much later arrive at workflows anyone can rely on. Freeform experimentation is a legitimate and necessary phase. It is a terrible permanent destination. Sooner or later the question has to shift from "What can individuals do with this?" to "How should the organization do this?"

That shift is where freeform gives way to the other two modes.

Transactional Coordination: Keeping AI on Task

For the past three decades, most of our interactions with business software have been transactional. The software exists to do a specific job, by a specific process, producing a specific output. A salesperson managing customer data uses the CRM. A CFO allocating budget uses the ERP. Someone building a deck uses PowerPoint. No experimentation, no reinventing the wheel—a tool configured to support a known process and deliver a consistent result at a predictable cost.

We tend to associate generative AI with the opposite of all that: open-ended chat, creative sprawl, surprise. But increasingly AI is being embedded into

exactly these structured workflows, the "vertical" applications built for one function—where it's responsible for just a few steps inside a larger predefined process. An expense tool that categorizes a line item by reading its description (airline name plus flight number equals "travel"). A proposal tool that pulls language from past winning bids into a new estimate for an existing client. Done well, this captures the best of both worlds: the creativity and tolerance for messy input we want from AI, wrapped in the consistency and process-discipline we want from traditional software.

We recently helped a construction company evaluate AI tools for the "takeoff" process, where estimators read blueprints and compute the materials and labor a project will require. It's painstaking work, and the stakes are real—a single footnote on page 36 can swing a cost estimate by hundreds of thousands of dollars. AI can't handle that kind of review 100% reliably end to end quite yet, but tools like Autodesk and Procore have added AI features which scan the blueprint, count every window and door, and highlight them for the estimator—saving hours of manual markup.

Notice what nobody in that engagement wanted? Nobody asked the estimators to "unlock new ways of thinking about takeoffs." Nobody expected the AI to draft a follow-up email to the client or muse on the relative merits of one window over another. The goal was narrow and known: fewer errors, less time, on a well-defined task. The organization already knew what success looked like. The AI just helped them get there faster.

That clarity is the strength of transactional coordination: when a task is this well-bounded, you can measure its cost, its cycle time, and its error rate precisely, and tune the system the way you'd tune a production line. But transactional coordination has two characteristic failure modes, and they sit on opposite sides of the same road.

The first is building something so constrained you'd have been better off with ordinary software. If your "AI-powered" workflow is really just a chain of rigid templates with no actual judgment exercised anywhere, you've added cost, latency, and a new failure surface in exchange for nothing. You took a deterministic problem and solved it with a probabilistic tool.

The second is the mirror image: building something users hate, so they route around it. They quietly pull out their phones, do the task in a personal chatbot account, and paste the result back in. Now you have shadow AI—unsupervised, ungoverned, invisible to management, while the official tool you paid for becomes shelfware. (We'll see this dynamic again in the next chapter,

with a corporate lawyer who abandons a $50,000 contract-analysis tool for ChatGPT because she can better adapt it to her needs. Whether you call that a failed implementation or successful adoption depends entirely on where you're standing—but either way, it's transactional coordination breaking down.)

The art of transactional coordination is staying in the narrow band between those two ditches: enough structure to be consistent and cheap, enough intelligence and usability that people don't defect.

Systemic Coordination: AI on a Mission

If transactional coordination is about embedding AI into a workflow, systemic coordination is about embedding AI into the organization itself.
Most discussions of AI never leave the level of "a human uses a tool to do a task." Even organizations running genuinely sophisticated AI tend to picture it as a drawer full of separate gadgets—one AI for HR, another for finance, a third for resource planning—each minding its own workflow. Systemic coordination dissolves those boundaries.

Here's a real example, lightly disguised. Imagine the set of systems supporting maintenance at a factory. One helps maintenance managers build schedules. A second monitors equipment-sensor data and flags likely failures. A third answers technicians' questions in the field about processes and machines. A fourth screens job applicants for technical aptitude and trains the ones who are hired on the company's SOPs.

Now imagine that, instead of presenting as four separate applications, all of these surfaced to users as a single persona. We'll call him Roberto.

It sounds like a gimmick, and at one level it is. But the shared persona does real work: it aligns the systems around a shared purpose. From the business's point of view, all four serve one mission—keeping the factory on its manufacturer-recommended maintenance schedule so that expensive things don't break. Roberto helps the manager plan the work. Roberto assists the technician on the floor. Roberto watches the equipment's vital signs. Roberto helps recruit and train the next technician. Underneath, it's a tangle of different models, prompts, databases, and reporting tools—but to the people using it, the complexity is invisible. They aren't operating a suite of applications. They're in an ongoing relationship with a persistent colleague who happens not to be human.

And from the organization's vantage point, something stranger and more interesting is happening than "people using tools." What's emerging is a kind of organizational chimera—a hybrid actor made of humans, software, and AI systems operating toward a shared goal. Initiative passes back and forth depending on who knows what, and when. The predictive system spots a failure before any human notices. A technician turns to Roberto for a diagnosis, or just to look up a spec. No single participant controls every step, yet from the outside the chimera behaves like one actor with a coherent purpose.

This is the mode where the deepest gains live—and also the mode that opens a Pandora's box of governance questions, most of them not technical at all.

Take the social-services systems we built for another client. Some of the governance questions answered themselves: obviously an AI coach assisting one client must never see another client's information. But others had no clean answer. Should the AI coach be able to read the human caseworker's private notes? Should caseworkers be allowed to read full transcripts of AI-client conversations, even when the client believed they were speaking to the AI in confidence? Should something disclosed in one context automatically surface in another?

These are not engineering decisions dressed up as ethics. They are questions about authority, memory, privacy, accountability, and trust—the exact questions humans have been negotiating among themselves for centuries. The novelty is only that one of the parties to the negotiation is now a machine. And as I'll argue throughout the rest of this book, no model can answer them for you. They are governance choices, and if your organization already makes such choices well, systemic AI will allow that discipline to pay off. If it has dodged them for years, systemic AI will simply expose those failures more quickly.

Choosing the Right Mode (and Catching the Wrong One)

I'm getting tired of pundits declaring how things will be "in the age of AI." But if there's a single managerial skill organizations need to cultivate "in the age of AI," it's the ability to look at a piece of work and ask: which kind of coordination does this actually want?

Much of the confusion in the market comes from people evaluating every AI initiative through a single lens. The "give everyone ChatGPT" crowd hopes individual experimentation will somehow resolve what are really process and organizational problems—they try to solve everything with freeform.

The workflow-automation crowd does the opposite, forcing inherently open-ended work into rigid pipelines, or treating a systemic challenge as a pile of disconnected tools. And the agentic-AI enthusiasts see every use case as a stepping stone toward full autonomy, even where autonomy is neither practical nor wanted.

The reality is that some work should stay freeform—the novel, the one-off, the genuinely creative. Some work is best kept transactional—the high-volume, well-defined, consistency-critical. And the highest-value applications tend to be systemic. But when AI implementation fails, the cause is usually mode-confusion. Some organizations stay stuck forever in freeform, never systematizing what they could. Others bolt AI onto a workflow nobody had bothered to define, with no supporting infrastructure underneath it. And others adopt a systemic capability but refuse to let it change who does the work, resisting the restructuring that would have made it pay off. Each failure is, at bottom, a piece of work trapped in the wrong mode.

Just as important is noticing when work migrates between modes without anyone deciding it should. This is subtle and it is everywhere. Take a support-ticket assistant that starts life as a tidy, transactional task: read the customer's latest message, draft a reply. It works. Then someone asks for one reasonable-sounding addition—have it glance at the customer's earlier tickets, so replies don't ignore prior context. Sensible. But "earlier tickets" has no natural stopping point, so the tool starts pulling the customer's entire history, surfacing a billing complaint from two years ago, apologizing for an outage that was fixed the same afternoon, contradicting a promise a human agent made last week. That is a transactional task trying to become a systemic one, without the governance, monitoring, or design rework a systemic capability requires. The drift feels like an improvement at every individual step. The sum is a mess.

When it comes to actually implementing all this, there's no single correct architecture. Some organizations will buy turnkey products with AI baked in. Some will lean on the AI features of platforms they already run, like Copilot or Gemini. Others will stitch together their own models, workflows, and applications function by function. None of these is inherently right or

wrong. What matters is whether the organization has clarity about what each human-AI interaction is supposed to accomplish, which tools can best support it, and how the result will be measured.

Organizations love to frame AI adoption as a procurement problem: which tools should we buy? It is really a management problem: how should all these different minds work together? The technology grabs the headlines, but for the first time in history we are coordinating not just people, but multiple kinds of intelligence operating side by side. The pyramid-builders of the AI era won't be the organizations with the most powerful models. They'll be the ones with the most coherent operating model for putting humans and machines to work on the same mission.

Chapter 4

Why Smart Humans + Smart Robots Do Dumb Things: Diagnosing AI Implementation Failures

"It's silly, no? / When a rocket ship explodes / and everybody still wants to fly."
—Prince, Sign O' the Times

***

On January 28, 1986, I stayed home from school with an upset stomach. I remember lying on the sofa in my Star Trek pajamas as my mother flipped through channels on the television,eventually landing on a broadcast of that day's space shuttle launch

"I'll be downstairs if you need anything," she said, leaving me to watch the countdown.

Even at that age, I had watched enough broadcasts and recordings of rocket launches that it all seemed routine. The countdown finished, massive clouds of smoke billowed from the launch pad as the skyscraper-sized boosters moved upwards, their exhaust nozzles blazing. Then, for a few minutes, the space shuttle just kept rising in the sky, trailing clouds of exhaust until, like some cruel magic trick, it was suddenly replaced by a fireball.

The television announcer fumbled for words:

"My God… a giant explosion… this is not standard procedure… this is not something that was planned…"

It took me a few minutes to process what I'd witnessed. At some point my mother ran in because I must have been screaming. Back at my school, they stopped classes so the teachers could talk to the other kids about what happened. In the days, weeks, and months that followed, the public and the government demanded answers about what went wrong.

The mechanical failure was straightforward: the unusually cold weather caused a rubber "O-ring" seal in the right rocket booster to lose its sealing ability, allowing superheated gas to leak through and ignite the external fuel tank. However, the organizational failures were more complex and significant.

NASA, the space agency, knew the seals didn't quite perform to specification, but because prior flights survived, they treated it as a minor, tolerable defect rather than a critical flaw. Engineers from the booster manufacturer warned NASA not to launch at temperatures below 11.7°C (53°F). However, not everyone involved was aware of these warnings. In the meantime, the agency was under tremendous pressure to maintain a busy launch schedule to justify the commercial viability of the space shuttle program. This led agency middle managers to pressure the rocket company's executives to sign off on the safety checklist, despite the engineers' objections, so the launch could proceed as planned.

Most of these details emerged long after the story had faded from the media and public memory. Over time, the public remembers explosions and forgets root causes.

Yet the larger story of the Challenger disaster is less about the horror and tragedy of that day in 1986 than about how the government and spaceflight industry responded in the decades that followed. Rather than abandoning space exploration, the United States invested billions to address immediate safety failures and reform its approach to human spaceflight. The disaster also accelerated a gradual shift away from the overly complex and expensive Space Shuttle model and back toward simpler launch systems, helping create the conditions that later enabled companies such as SpaceX and the modern commercial space industry. Most importantly, Challenger demonstrated that while technologies can fail catastrophically, organizational culture, incentives, and decision-making processes are often the deeper causes of disaster.

Modern AI is still a young technology compared with spaceflight, yet it has already produced several high-profile failures. Zillow, a real estate technology company, lost more than $500 million when its AI-driven home-buying program systematically overpaid for properties in rapidly changing markets. Amazon abandoned an experimental hiring algorithm after discovering it disadvantaged female applicants. Fast-food chains including McDonald's and Taco Bell attracted public ridicule when AI drive-through ordering systems became trapped in endless question loops or crashed after pranksters manipulated them with absurd requests, including an infamous attempt to order 18,000 cups of water.

However, those are all edge cases in that:

- Most of the failures we know about are old, involving previous generations of AI technology
- They happened in public
- They were obvious enough for people to notice
- They were outrageous enough for news outlets to cover

The majority of AI failures possess none of those qualities. When an implementation fails at an ordinary mid-market company, the executives who championed it are embarrassed, employees are inconvenienced, pilot programs stall, budgets are cut, teams quietly return to spreadsheets, and the episode disappears without notice. The most consequential AI failures are often not spectacular disasters but mundane organizational disappointments that never become news.

Researching organizational failures has always been challenging, since people generally don't want to discuss them (or can't discuss them per company policy or NDAs). When I asked my own contacts at large organizations about any failures or frustrations their teams encountered, they all politely changed the subject or flat-out declined. And that's entirely fair since I probably wouldn't share any stories about my own consulting clients' struggles, even anonymously, if I knew about them.

This leaves us with more general industry surveys, like those conducted by MIT, RAND, and Gartner in 2025 and 2026, which place the failure rate of AI implementations anywhere from 75% to 95%. But while these studies are widely cited by critics of AI, they offer limited insight into exactly why projects failed. Specifically:

- The MIT and RAND studies were both based on interviews with executives, not a thorough examination of the specific tools and systems involved or how they were developed and implemented.
- The studies mainly gauged sentiment and perception, not hard performance data (e.g., the MIT study quotes one executive as saying, "The hype on LinkedIn says everything has changed, but in our operations, nothing fundamental has shifted. We're processing some contracts faster, but that's all that has changed.") [6]
- The findings were mostly presented in aggregate, with just a few anonymous quotes here and there, making it hard to gain anything but surface-level insights from the failures.

None of this is to deny that the failure rate is massive (that part I believe). It is simply to point out that these studies offer little in the way of hard data or detailed case studies from which to derive lessons.

Yet, despite the limited evidence available, we can still draw some general conclusions about why AI implementations fail. These conclusions are based on a combination of documented incidents, industry surveys, anonymous employee confessions posted on Reddit, and my own observations from consulting work. While these sources involve different organizations, industries, and AI systems, the underlying behavior patterns and failure modes are remarkably consistent.

Lack of a Clear Business Objective

The following statement is so obvious that I feel stupid typing it: AI is a technology, not a business objective.

Yes, AI can summarize documents, generate images, tag database records, retrieve knowledge buried in old document libraries, brainstorm marketing strategies, and automate workflow steps that previously required human intervention. But unless those capabilities help your organization generate revenue, reduce costs, improve quality, reduce risk, or advance its mission at a reasonable cost, all that AI activity is just an expensive distraction.

You'd think all that would go without saying, yet history shows that once the hype around a new technology exceeds a certain threshold, otherwise sensible organizations start asking the wrong questions: instead of "What outcome are we trying to improve?" they ask, "How can we use the shiny new toy?"

It happened with computers. It happened with the web. It happened with mobile and the cloud. Now it is happening with AI to a degree where, if the bubble pops, the entire global economy might implode.

Many AI technology companies and pundits, including some I otherwise respect, are acting as enablers. "Just give your staff access to AI and let them experiment," they say in Substacks and Ted talks, "and trust that valuable use cases will emerge."

---

[6] Challapally et al., The GenAI Divide, 8.

There is some truth in this: freeform experimentation and creative play can reveal opportunities leadership would never identify through a top-down planning process. But some organizations have taken this concept to the illogical extreme.

Meta and Amazon created incentive programs based on how many tokens (units of AI compute) [7] their developers consumed each month, even going so far as to set up leaderboards for who could use the most compute. In the end, Meta was forced to pull the plug as developers were having AI agents perform unnecessary or redundant tasks just to drive up their scores. Another tech giant, Uber, burned through its entire annual AI budget in just four months without any corresponding productivity gains.

Yes, this actually happened (go look it up). And yes, it's as dumb as it sounds.

Token consumption measures AI activity. It does not measure business value. If Amazon suggested rewarding its delivery drivers for burning more gasoline, everyone would have called them insane. Yet somehow that kind of logic gets a pass with AI, especially when an organization can't think beyond a "Freeform" usage mode, leading them to mistake individual unstructured experimentation for a coherent AI strategy.

But if measuring activity doesn't work, what new metrics do organizations need to invent to measure the ROI for AI?

The beauty of it is that, most of the time, you do not need to invent new metrics at all.

In our own work, one of the clearest examples of this came from a factory where we helped design an AI system to assist preventive maintenance technicians. It would have been easy to define success in AI-flavored terms: number of users, number of prompts, number of work orders touched by AI, or number of recommendations generated.

Instead, we arrived at a much more concrete measure: hydraulic oil.

---

[7] Mark Minevich and Simon Ninan, "Most Businesses Are Measuring AI Wrong, and It's Costing Them," Fast Company, June 24, 2026, https://www.fastcompany.com/91555955 /most-businesses-are-measuring-ai-wrong-and-its-costing-them-ai-tokens-strategy.

The factory had a recurring problem. When technicians encountered leaking hydraulic systems, they often refilled the oil rather than identifying and repairing the leak. From a narrow task perspective, this was understandable. The machine needed oil, the technician added oil, production continued, and everyone moved on. But from a maintenance perspective, the behavior was expensive and counterproductive. The underlying issue remained unresolved, oil inventories depleted faster than they should, and the same problems reappeared.

The goal of the AI preventive maintenance copilot was not to "increase AI adoption" among technicians. It was to help technicians diagnose recurring equipment issues, provide guidance at the moment of need, and hold them accountable for explaining their daily maintenance activity in enough detail that managers could tell whether problems were actually being fixed.

If the system worked, we expected to see it in the hydraulic oil stocks: less oil disappearing into the same leaking systems, more leaks repaired, fewer recurring top-offs, and better maintenance behavior.

Milliliters, not tokens.

Failure to Systematize

One of the most transformative aspects of generative AI is how it can accelerate the work of experts and elevate the capabilities of non-experts. A salesperson can get answers to basic technical questions without having to bother an engineer, an attorney can review a stack of documents without having to pull in a paralegal, and an administrative assistant can build a working automation without ever filing a ticket with IT- often just using general purpose chatbots.

Most of the time, this is a good thing: when you give the people closest to a problem the tools to solve it directly, it can unlock a level of creativity and agility that rarely happens when people need to go through layers of bureaucracy.

That said, there are times when handing power tools to inexpert users leads to accidents. What once required a developer (and, by extension, a ticket, a spec, a security review, and a release process) can now be assembled by a motivated marketing intern before lunch. What once required a subject-matter expert's sign-off now gets drafted, finalized, and sent while that

expert is still in another meeting. The work gets done, but without the checks and balances that organizations traditionally relied on to ensure quality, consistency, and adherence to process.

One visible symptom of this breakdown is the rise of what researchers have called "workslop": AI-generated work that looks polished but is largely devoid of value. A study published in Harvard Business Review of 1,150 U.S. workers found 40% had received substandard work products, reports and communications from peers that were clearly generated by AI. [8] Likewise, 18% of managers said they received "slop" from their subordinates, and 16% of subordinates had received slop from managers.

Sometimes, leadership is to blame for this, either by failing to communicate and enforce a reasonable and coherent policy on the use of general purpose AI chatbots, or by failing to provide approved AI tools that meet their users' real-world needs. One corporate lawyer interviewed for the MIT study described this tension clearly. Her organization had spent approximately $50,000 on a specialized contract analysis tool, yet staff routinely reverted to ChatGPT for day-to-day work:

> ***"Our purchased AI tool provided rigid summaries with limited customization options. With ChatGPT, I can guide the conversation and iterate until I get exactly what I need. The fundamental quality difference is noticeable. ChatGPT consistently produces better outputs, even though our vendor claims to use the same underlying technology."***

Whether this represents a failed contract analysis implementation or successful ChatGPT adoption depends entirely on perspective (viewed through our 3-level framework, we'd call it a Transactional coordination failure resulting in a retreat to Freeform shadow AI).

One might reasonably ask "What's the harm? If people prefer general-purpose chatbots, and demonstrate they can use them responsibly, why not let them?"

In some cases, there really isn't any harm and using a general purpose chatbot is the right move. However, in other cases, unsystematic use of general purpose AI creates serious risks.

---

[8] Kate Niederhoffer et al., "AI-Generated 'Workslop' Is Destroying Productivity," Harvard Business Review, September 22, 2025, https://hbr.org/2025/09/ai-generated-workslop-is-destroying-productivity.

Most commercial AI models prioritize fast answers to a user's immediate questions, not the rigorous discovery a human professional would conduct before a consequential decision like approving a loan, redesigning a factory's layout, or making a mental health diagnosis. A seasoned expert knows what they don't yet know and goes looking for it: and while it's possible to instill that rigor into a purpose-built AI tool system, a general-purpose chatbot is unlikely to ask more than one or two clarifying questions before offering a recommendation.

We've seen this play out with clients and within our own organization. In one case, a client's non-technical salesperson had ChatGPT generate a proposal that fundamentally misread the customer's problem and misrepresented the capabilities of the company's products: not because the AI was "stupid," but because it took the salesperson's description at face value and filled in gaps, rather than asking for supporting evidence or product specifications.

Even when individuals use AI effectively, overreliance on ad hoc AI solutions creates the risk of process fragmentation. One of our clients encouraged staff to use Gemini for producing standard reports, resulting in individual reports that looked professionally and attractively formatted, but varied wildly from one employee to another and one week to the next, like salads assembled from entirely different recipes.

There is also a quieter cost: when AI use remains ad hoc, it's harder to share best practices and improve processes. One person might develop an excellent prompt for a common work task while their peers either avoid AI entirely or produce lower-quality outputs through trial and error. Unless successful experiments are identified, tested, refined, and turned into SOPs, AI use remains a laboratory rather than a factory.

Again, none of this is to say that general purpose AI chatbots are bad: some AI use should remain freeform and flexible. Rather, the true villain is the lack of a coherent operating model and clearly communicated policies, guidelines, and rationales for what should be done with general purpose chatbots and what should be done either unassisted or with approved, purpose-built tools.

Absence of Supporting Infrastructure

Since the 1990s, practitioners of "knowledge management" have argued that making the right information available to workers at the moment of need is often more valuable than a hundred hours of training delivered in

advance. Rather than expecting employees to memorize everything they might someday need to know, organizations should design systems that help people find, interpret, and apply the right knowledge when the work actually demands it.

I have always loved that idea… at least in theory.

In practice, most knowledge management systems that start as elegantly architected temples of institutional knowledge eventually degenerate into SharePoint folders full of redundant files, outdated policies, half-finished drafts, and documents named "Final_Final_Updated_v7_REALLY_FINAL." Customer relationship management systems accumulate ten different entries for the same client. Internal wikis become archaeological sites. Search boxes either turn up nothing or vomit forth seven pages of irrelevant results before finally returning a relevant policy document, which might or might not be up to date.

I am not bashing knowledge management practitioners. Quite the opposite. My company used to have an entire knowledge management practice that represented 20 to 30 percent of our revenue, depending on the year. But while we genuinely believed in the value of KM, and enjoyed helping organizations get their institutional knowledge in order, we eventually grew tired of fighting the inevitable tide of decay as knowledge bases fell out of date, file naming and tagging conventions slipped, owners moved on, KM budget got cut, and no one had time to maintain the system everyone had agreed was so important.

Case in point: my company was once approached by a client that distributed auto parts to major Detroit car manufacturers. They explained that one of the Big Five consulting firms had previously tried and failed to develop a system for reconciling the company's various supplier catalogs with the data formats required by its customers' ordering systems. The problem was messy, technical, and deeply unglamorous: different suppliers described similar parts in different ways, while customers required information to be structured according to their own ordering conventions.

After considerable work, we developed a taxonomy that could square the circle. It gave the company a way to translate between supplier catalogs and customer requirements with far greater consistency.

The client liked the solution. Then came the inevitable question: "Great. Now who's going to maintain this?"

We told them the truth: they would need to hire someone. In fact, they would probably need to hire two people.

Their response was immediate: "Ain't gonna happen."

And with that, the solution effectively died.

That story captures a kind of organizational debt that has been building for decades. Companies have long wanted clean data, reliable knowledge repositories, well-maintained taxonomies, current documentation, and clearly defined workflows. They have simply not wanted to pay for the human effort required to create and maintain them.

Organizations have also spent decades looking for shortcuts around this work. For a while, enterprise search was supposed to be that shortcut. Companies like Microsoft and Google sold the promise that if employees could search across the intranet, shared drives, email, and document libraries, organizations would not need to be quite so disciplined about structure, ownership, metadata, version control, or archiving.

But while search helped, it did not solve the underlying problem. It made messy repositories easier to query. It did not make them less messy. A search engine could surface five versions of a policy, but it could not always tell employees which one was current, which one had been superseded, and which one should have been deleted three years earlier by someone who has since left the company.

Now, many organizations are hoping AI will provide the quick fix for KM that search couldn't deliver, allowing employees to simply ask an AI agent questions about the latest data security policy and magically receive a plain-language answer, without having to maintain all those document repositories, systems of record, taxonomies, and workflows.

Alas, this gets the relationship exactly backward. In reality, if you connect Microsoft Copilot, Gemini, or another AI agent to a document library full of redundant, outdated, trivial, contradictory, and poorly named documents, the result will not be organizational intelligence. It will be faster access to organizational confusion. The AI will retrieve from the mess, summarize

the mess, and, because it is AI, make the mess sound polished enough that someone may mistake it for truth.

The same is true for systems of record and workflows. AI agents can be excellent interfaces for helping people retrieve information, make updates, escalate issues, draft responses, and complete workflow steps that once required human intervention. But they depend on the quality of the systems underneath them. If the customer database contains duplicates, the agent inherits the duplicates. If the product database contains outdated specifications, the agent retrieves outdated specifications. If the workflow is undocumented, inconsistent, or different in every region, the agent cannot magically infer the one true process from vibes and corporate folklore.

Knowledge management, systems of record, and workflow discipline are not optional accessories to AI implementation. They are prerequisites: one might say AI needs knowledge management more than knowledge management needs AI.

Yet organizations still hope AI will provide a cheat code to render KM unnecessary.

Our team once designed an AI agent that could read and update inventory data in a client's ERP system while also retrieving relevant information from their product database. The demo worked well. Users could ask natural-language questions, retrieve inventory details, and make certain updates without navigating the ERP interface directly.

After seeing the demo, the client said, "This is great. Do you think we could use the AI agent instead of the ERP for tracking inventory? Our ERP rollout has been falling behind schedule."

We explained the distinction this way: think of the AI agent as more like a banker than a bank ledger. I would trust a banker to look up my account balance, answer questions, and help move money between accounts. But I wouldn't want the banker's notebook to become the official record of my balance. The ERP is the bank ledger, the source of truth. The agent is simply an interface to it.

The client understood, but you could tell that a little of the "AI magic" had faded, replaced by the far less glamorous reality of enterprise data architecture.

Lack of Internal Capacity

Having done both, I can say there is a big difference between working in a support function and working in an organization's core function. Being a nurse at a school is very different from managing the nursing unit in a hospital emergency room, just as overseeing IT at a construction company is very different from being a solution engineer at Amazon Web Services.

For professionals, the biggest upside of working in a support function is that you can apply the same basic skills and methodologies across many different industries. While some industry-specific domain knowledge is always required, it's not uncommon for someone to work as a financial analyst at a chemical plant for five years, then somehow find themselves doing the same job at a clothing retailer for the next five years. In my own career, I've worked in the corporate training department of a company that sold security cameras and ran the training department of a company that made healthcare management software, and while I've never worked in the food industry, I'm sure I could competently run the training function of a grocery chain if given 4 to 8 weeks to learn their business. That's not boasting: that's just how it is for most professionals above a certain level of seniority.

Support functions also tend to prioritize reliability over experimentation. While it's important for leaders in support departments to keep current, their mission is to enable the organization's core work, not to operate at the cutting edge of their respective disciplines. A construction company wants excellent IT support, not an IT department conducting advanced computer science research.

If you're reading this book I assume you find AI technology interesting—perhaps even fascinating. However, we need to acknowledge that most organizations implementing AI systems won't be AI companies. They'll be manufacturers, healthcare providers, financial institutions, educational institutions, and professional services firms for whom AI is ultimately a supporting function. Yet unlike accounting, human resources, information technology, or learning and development, there is not yet a large and mature labor market of experienced AI practitioners who can move easily between organizations and apply proven methods.

All this is to say two things:

- Most organizations lack internal AI capacity.
- That shortage reflects the state of the labor market as much as an organization's sophistication.

Now, when we talk about a lack of internal capacity, we mean two different things.

The first is technical capacity: having people with enough technical knowledge to evaluate tools, build solutions, oversee vendors, and / or support systems after implementation.

The second is translation capacity: having people who understand both the business and the technology well enough to connect the two. Organizations rarely fail because they lack access to AI tools (throw a rock and you'll hit a company selling "AI for chiropractors" or "AI for waste management"). They fail because nobody inside the organization can effectively translate between business needs, operational realities, and technological capabilities.

Workers with AI technical capacity are in short supply. Workers with domain-specific translation capacity are rarer still.

So, does this mean there's no hope for the average veterinary clinic looking to implement AI solutions? Hardly.

The MIT State of AI in Business 2025 study found that successful AI implementations typically combined external expertise with internal "power users" who already used AI tools extensively on their own: in other words, a combination of technical and translation capacity. [9] These projects were more than twice as likely to move beyond the pilot stage and achieved significantly higher employee adoption rates while reaching value more quickly and at lower cost.

In our own work, we've seen the difference between projects where these capacities were present and projects where one or both were absent.

---

[9] Challapally et al., The GenAI Divide.

- We've seen restaurant chains hire machine learning experts from major Silicon Valley tech companies, only to deliver solutions that failed to account for the chaos and congestion of a busy kitchen or dining room.
- We've seen projects led by enthusiastic internal AI advocates who understood how the technology could transform customer self-service, but who lacked the technical expertise to build and maintain the data pipelines their ideas demanded.
- At the other extreme, we've seen organizations assign AI initiatives to domain experts whose skepticism became a self-fulfilling prophecy, as every minor error produced by an early prototype was treated as a fatal flaw, and every implementation challenge became evidence that the project should be scrapped.

However, we've also been fortunate enough to work on projects alongside client-side "power users" who understood the business problem and grasped the technology well enough to participate meaningfully in design, testing, and iteration. And that's often the difference between a system that delivers a modest 30 percent productivity improvement and one that fundamentally changes how work gets done.

Failure to Maintain / Iterate Systems

The most pernicious myth about AI is that "AI" is a glowing blob of intelligence that "learns" over time, and it somehow gets wiser, more accurate, and more capable the more users interact with it.

The reality is far more complicated and far less magical.

What most users refer to as "AI" is really a complex computer system consisting of multiple components—models, data sources, prompts, tools, agentic harnesses—working together in concert. And while it is possible to set up feedback loops that help the AI system become more context-aware and effective over time, this doesn't happen automatically, and performance is just as likely to degrade as improve without quality assurance testing, monitoring, maintenance, and periodic upgrades.

This means keeping its data sources current and making sure the system is retrieving the latest information rather than old policies and discarded drafts. It means listening to user feedback and making adjustments. It means replacing models as new ones are developed and old ones are deprecated, then running it through a quick series of tests to make sure the new version

doesn't introduce any quirks. It means updating instructions when processes and systems change, and doing spot checks of outputs to make sure quality isn't degrading at scale.

This is where even those organizations who successfully launch an AI system get into trouble. They assume AI maintenance isn't necessary or is somehow different from ordinary software maintenance, when the real problem is that it requires all the usual maintenance and more.

One of my favorite ways to help clients dispel wishful AI thinking is to replace the word "AI" with "computer" in common claims, making the nonsense easier to detect.

- "We deployed the computer system and assumed it would improve automatically."
- "We fed the computer system outdated files and were surprised when it gave outdated answers."
- "We launched the computer system without deciding who owned its performance."
- "We assumed the computer system could handle our use case at scale because the demo was really impressive."

Nobody would say these things about an ordinary software system with a straight face. Yet the same logic is often tolerated with AI because the technology carries a fog of mystery around it. That fog is useful to vendors, consultants, and pundits. It is less useful to the organizations paying the invoices.

AI implementation is not one and done. Once it becomes part of your infrastructure, you either maintain it or you let it crumble.

Pretending ROI Is Mysterious

Another common reason AI implementations fail to deliver measurable ROI is that organizations needlessly overcomplicate measurement.

When rolling out AI, most companies tend to measure activity. They track how many prompts employees submit, how often they log into a tool, how many documents an AI system summarizes, or how many times an agent is invoked. But while those numbers may be useful operational signals, they do not in themselves represent ROI.

In most other cases, the measures of value are familiar, even mundane: lower cost, shorter cycle time, higher quality, reduced risk, increased throughput, improved customer service, faster onboarding, or better use of scarce expertise.

So here's a radical proposition: why not measure AI by those same numbers?

If an AI system helps perform work that humans used to perform, then its costs should be compared directly to the cost of the human labor it replaces, accelerates, and / or enhances.

If a task that once required five hours of first-pass human review now requires twenty minutes of AI processing, thirty minutes of human setup, and one hour of expert review, then our budgets should reflect that, with the AI token costs accounted for alongside the human labor hours. If employees receive AI seat licenses, those licenses should be treated as part of their fully burdened cost, just like software, equipment, benefits, office space, and anything else required for them to do their jobs.

This is not metaphysics. It is math:

Previous process = human labor cost + cycle time + error rate + review cost + overhead.

AI-assisted process = AI cost + human setup cost + human review cost + cycle time + error rate + AI-specific infrastructure and maintenance cost.

Note that labor costs and token costs are not the only components here: cycle time is perhaps the most underrated component of AI ROI.

We saw this with a client in the financial services industry that needed to review a large body of new regulations and determine which provisions were relevant to its business. The material ran to hundreds of pages at the outset and likely exceeded a thousand pages by the time the project was complete. Under the existing process, the review would have taken approximately seven months. With the AI-assisted workflow, the work could be completed in about four weeks.

The client was happy about the cost savings, but that was not what excited them most. The real value was speed. A seven-month review cycle meant the organization would spend more than half a year waiting to understand the implications of new regulations. A four-week review cycle meant it could

begin responding while the information was still operationally useful. Meanwhile, in terms of error rates, the AI system's work was judged acceptable 92% of the time, compared with roughly 78 to 84% for human reviewers.

To give another example from our own work, we were able to substantially accelerate the onboarding process for new software developers with AI.

In the past, we would expect a new developer to make their first useful code commit on the afternoon of Day 3. Today, it is not unreasonable to expect a useful code commit by the end of Day 1. The difference is that new developers can now ask Claude Code to direct them to where a particular feature is addressed in the code base, versus having to read through it themselves. The AI system does not make them an expert on the codebase overnight: it just provides a faster way for developers to orient themselves, ask better questions, and become useful sooner.

None of these figures requires a special theory of AI ROI. They required comparing the old process to the new one: time, cost, quality, and human review burden.

Granted, this approach is bound to make several groups uncomfortable:

- It annoys AI evangelists who want adoption without accountability.
- It annoys executives whose organizations never had clear baselines in the first place.
- It annoys managers who could not tell you how much capacity their teams actually have or where freed-up time would go.
- It annoys workers who do not want their performance compared against AI-assisted workflows (nor, for that matter, against any kind of benchmark).
- It especially annoys people who would prefer to measure enthusiasm, usage, and "transformation" instead of cost, time, quality, and output.

While organizations will want to take these sensitivities into account for the sake of change management, it's never too soon for organizations to start tracking the hard numbers.

When building AI systems for clients, we will typically benchmark the first few runs at an extremely granular level: noting which models were used, how many messages or tool calls were exchanged, how much each of those operations cost, how much human setup or review was needed, and how the output compared with the prior human process. That early benchmarking often reveals opportunities to reduce costs or improve consistency by making the workflow more structured.

Admittedly, some AI proponents might find this stifling: holding a nascent technology to rigid expectations when they should be encouraging experimentation. But while there's nothing wrong with giving teams room to explore AI tools, test workflows, build prototypes, and learn what is possible, exploration needs to be recognized as such.

Resistance to Restructuring

The deepest AI gains rarely come from helping the same people do the same work slightly faster. They come from rethinking the work itself. That is also why these gains are so difficult to capture. A tool that saves an expert 30% of their time may be useful, but a tool that allows a junior employee to perform the first pass, escalate only the ambiguous cases, and reserve expert attention for true judgment work may transform the economics of an entire process. The technology may be impressive, but the uncomfortable question is organizational: if AI changes who is capable of doing the work, are leaders willing to change who actually does it?

To give a practical example, here's a story from one of our client engagements (the industry specifics have been changed to protect the client, but the structure of the problem remains the same).

We worked with an engineering firm whose senior engineers were spending an inordinate amount of time reviewing parts in vendor catalogs, only to discover that most would ultimately fail to meet project requirements. We helped them create an AI agent that could scan an online catalog entry in seconds and, in comparison tests, reached the same conclusion as a senior engineer in 19 out of 20 cases. The remaining cases were typically outliers where even the engineers felt they would need to contact the supplier for clarification before making a final determination.

To maximize the time saved, we suggested delegating the initial review process to a junior engineer (or at least having one engineer perform the searches on behalf of the team) and then forwarding a shortlist of viable options to the senior engineers for final review. However, the client chose to have the senior engineers continue conducting the searches themselves, using the AI agent as an assistant.

In a three-week pilot, the result was a disappointing 30–50% reduction in search time, varying by engineer, along with widespread complaints that the tool was more trouble than it was worth.

This made the client's executive stakeholders nervous, but they acknowledged that the process was always meant to be iterative. Our team took the opportunity to analyze how different engineers were using the system, and found substantial variation in search behavior. Some engineers gave up quickly, while others continued searching long after they had identified several viable options, generating diminishing returns. Some performed a shallow review across many supplier catalogs, while others examined a small number of catalogs in much greater depth.

Based on these observations, we developed workflow guidelines that reflected the most effective search patterns. We also simplified portions of the interface by replacing common freeform interactions with button-driven actions such as "Scan Now" and "Summarize Now," reducing both user effort and token consumption. Finally, we aligned model selection with task complexity, assigning a more capable model to the evaluation step while using a smaller model for routine summarization and database updates.

The next month, we did a second pilot. This time, the searches were delegated to a junior team member who was still in graduate school, and we spent an afternoon training them extensively while collecting additional usability feedback. After two weeks, they achieved a 75–90% reduction in search time, depending on the complexity of the requirements.

Yet even with those results, some senior team members remained hesitant to hand off the task. For some, the idea of shifting responsibility from senior engineers to a junior colleague supported by AI felt like a diminishment of their role, even though they had plenty of other, higher-value work to do.

What the client resisted was the exact threshold between Transactional and Systemic coordination. They were happy to use AI to speed up a Transactional task, but ultimately unwilling to make the Systemic changes to their org chart required to actually unlock the technology's full economic value.

And this is just a small manifestation of a much broader challenge in AI adoption. Organizations often assume that the primary challenge is making the technology work. In practice, the larger challenge is redesigning work around new capabilities. AI creates the most value when companies reconsider who performs which tasks, how decisions are escalated, and where expertise is actually required. Taken to its logical conclusion, these changes can affect staffing mixes, job descriptions, hiring priorities, departmental missions, and organizational structures upon which many stakeholders have built their careers.

Our experience and conclusions in this case were consistent with the MIT State of AI in Business 2025 study, where one manufacturing operations executive said "Our hiring strategy prioritizes candidates who demonstrate AI tool proficiency. Recent graduates often exceed experienced professionals in this capability." [10] But while it's easy to demand AI literacy as a requirement for entry-level positions going forward, how many organizations are prepared to retrain, reorganize, and reallocate work across the employees they already have?

Conclusion

What happened to the Challenger and its crew in 1986 was a terrible disaster: yet abandoning spaceflight would have been even more disastrous when you consider how much society has benefitted from continuing to explore and commercialize space in the decades since.

This isn't to say that organizations have any obligation to push on with AI implementation for its own sake. If you're not seeing ROI, then sometimes the correct response is to stop, cut losses, and admit that a use case was poorly chosen, a vendor was oversold, a workflow was not ready, or the expected value was never there.

But there is also a danger in learning the wrong lessons from failure. A stalled chatbot, disappointing pilot, or embarrassing prototype does not prove that AI technology has no value. It may prove only that the organization lacked a clear objective, reliable infrastructure, internal capacity, maintenance discipline, honest measurement, or the willingness to restructure work around the new capability.

AI failure is rarely mysterious. It is usually diagnosable, if people know what to examine. And what can be diagnosed can usually be fixed, if we believe in the value and are prepared to put in the work.

---

[10] Challapally et al., The GenAI Divide.

Thank you

for reading these preview chapters.

To download the entire book visit www.parrotbox.ai/books
or email info@parrotbox.ai.

The complete book contains the following chapters:

Chapter 1: One Weird Trick: How AI Does What It Does

Chapter 2: Bots of Burden: From Artificial Intelligence to Real Work

Chapter 3: Freeform, Transactional, Systemic: Three Levels of Human-AI Coordination

Chapter 4: Why Smart Humans + Smart Robots Do Dumb Things: Diagnosing AI Implementation Failures

Chapter 5: Feeding the Beast: Why AI Is a Knowledge Architecture Problem, Not a Model Problem

Chapter 6: AI Is a Factory: Applying Lean Manufacturing Principles to AI Systems

Chapter 7: Big Brother Is Watching You (But Does He Truly See You?): From Reporting to Operational Awareness

Chapter 8: Taming the Terminator: Guardrails, Guidance, and Governance

Chapter 9: Roping and Riding: The Role of Humans on a Robot Ranch

Chapter 10: Digital Darwinism: How Organizations Adapt to Abundant Cognition