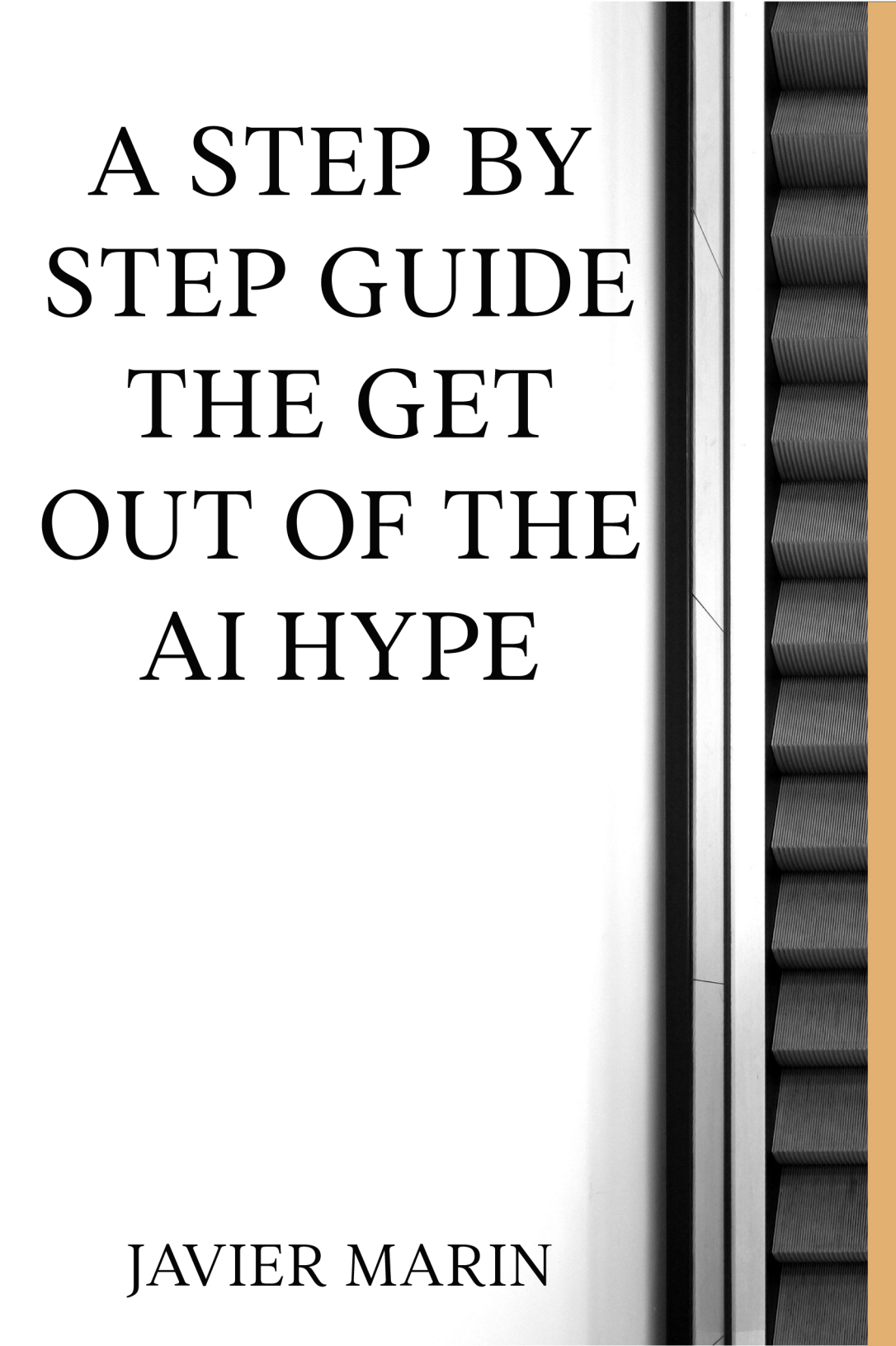
A STEP BY STEP GUIDE THE GET OUT OF THE AI HYPE



JAVIER MARIN

This work is based on an intense experience deploying AI systems for two years, and on a quantitative research of AI industry developments from January to June 2025. All data sources are cited and verifiable. The analysis represents the author's interpretation of publicly available information.

Dedicated to my beloved mother

Contents

Introduction

hype *[hahyp]*

verb (used with object)[1]
hyped, hyping

1. to stimulate, excite, or agitate (usually followed by *up*).

   She was hyped up at the thought of owning her own car.

2. to create interest in by flamboyant or dramatic methods; promote or publicize showily.

   a promoter who knows how to hype a prizefight.

3. to intensify (advertising, promotion, or publicity) by ingenious or questionable claims, methods, etc. (usually followed by *up*).

[1] `https://www.dictionary.com/browse/hype`

4. to trick; gull.

In late 2022, like many, I was genuinely engaged by the emergence of ChatGPT. I completely missed what was happening in the top AI labs. As it turns out, Large Language Models (LLMs) had been "cooking" for quite a few years. The signs were there, but I simply didn't recognize them. Even as a regular reader of distill.pub [2], I'd gone through articles like "*Visualizing memorization in RNNs*"[3]. The very issues and limitations of RNNs highlighted in that article – especially around long-term memorization and the vanishing gradient problem – were precisely what Transformer architectures were engineered to overcome. Today, distill.pub's editors, Chris Olah and Shan Carter, lead the Interpretability department at Anthropic.

Our team quickly began testing, and it became clear almost immediately that this technology would fundamentally alter many things. As a Data Scientist, I found myself concerned about a future I didn't yet fully grasp. The recent introduction of Copilot had already been a significant achievement for coding; swiftly, I saw many junior developers crafting elegant, highly efficient Python code. Just weeks earlier, their code had often appeared elementary, less polished. That alone was a leap forward. But then, the conversation truly shifted. Customers, colleagues,

[2] https://distill.pub//

[3] https://distill.pub/2019/memorization-in-rnns/

and many people around me began querying ChatGPT. Sometimes the questions were straightforward, sometimes quite specialized. And ChatGPT, more or less accurately (now we know it's called "hallucinations"), provided answers in a way that empowered users, giving them a sense of "I can do anything with this." A few months later, Microsoft's AutoGen arrived and triggered that 'this is the future' moment for many people. Some are still stuck there.

AI technology has advanced 10× over just two years. Yet, that initial sense of wonder, it hasn't changed much. While technology can redefine "its" world in milliseconds, our human understanding and acceptance might take years. This fundamental asymmetry – the velocity of technology versus the pace of our comprehension – is a key driver of the hype we're currently experiencing. A visionary from Silicon Valley might confidently assert that truly revolutionary products solve problems people weren't even aware they had. And that's probably true. However, we also need revolutionary products that address the problems we do know we have. Companies that have dedicated substantial resources to AI feel compelled to "agitate" the market with campaigns and promises that are sometimes far from reality. These factors combine to create the hype. As we'll explore, building trust is now a major challenge for many AI companies in the application layer – those demanding tokens from foundational models to solve specific problems.

Organizations getting real business value aren't the ones with the biggest AI budgets or the most sophisticated models. They're the ones that learned to separate signal from noise — and they follow a systematic approach to get there.

This book isn't an academic study, nor is it a how-to guide for developing AI projects. It's also not a deep dive into technical or business opinions. It's a practical guide that combines different perspective. The goal: help executives gain confidence in AI systems.

confidence *[kon-fi-duhns]*

noun[4]

1. full trust; belief in the powers, trustworthiness, or reliability of a person or thing.

   We have every confidence in their ability to succeed.

   Synonyms: dependence, reliance, faith

   Antonyms: mistrust

2. belief in oneself and one's powers or abilities; self-confidence; self-reliance; assurance.

   His lack of confidence defeated him.

3. certitude; assurance.

   He described the situation with such confidence that the audience believed him completely.

4. a secret that is confided or imparted trustfully.

   The friends exchanged many confidences over the years.

5. vote of confidence.

6. presumption; impudence.

   Her disdainful look crushed the confidence of the brash young man.

7. *Archaic.* something that gives confidence; ground of trust.

[4] `https://www.dictionary.com/browse/confidence`

01

How Big Is the Gap Between AI Promises and Reality?

In God we trust. All others must bring data
- W. Edwards Deming

In January 2025, something unprecedented happened in the AI industry. For the first time since the current boom began, we could actually measure the gap between what AI companies promised and what they delivered. The numbers were eye-opening—and they reveal patterns that every executive needs to understand.

This isn't another opinion piece about whether AI is over-hyped. This is six months of tracking what actually happened when AI companies had to deliver real results to real customers. What we found will change how you think about every AI proposal that lands on your desk.

A Simple Test

Let's start with something you can verify right now. Open LinkedIn and scroll through your feed. According to a comprehensive analysis by Originality.AI, 54% of all long-form content you're seeing was generated by artificial intelligence[1]. That represents a 189% surge since ChatGPT launched.

Engagement rates on AI-generated posts are 45% lower than human-written content. People seem to instinctively recognize and avoid content that feels artificial, even when they can't consciously identify why. But users keep shifting to AI-generated content anyway. It's easier and faster.

LinkedIn's own experience proves this point. Their AI writing assistant, launched with great fanfare, flopped spectacularly. CEO Ryan Roslansky admitted in March 2025 that the tool was "not as popular as I thought it

[1] `https://originality.ai/blog/ai-content-published-linkedin`

would be"[2]. The reason? Professionals didn't want to risk their reputation by posting content that might be recognized as artificial.

This creates a fundamental tension in AI content: you can create more of it, but you can't create more trust. The more AI-generated content floods platforms, the more valuable genuine human insight becomes. But we haven't arrived yet at this inflection point, and this makes AI content confusing.

The AI Startups Reality Check

Now let's look at what happened to the companies making the biggest AI promises. The numbers here are revealing. An analysis of AI startups founded in 2024 shows that 90% failed within their first year[3]. Industry analysts suggest 99% of AI startups will be dead by 2026[4]. The most dramatic example was Builder.ai, which raised $450 million from Microsoft and other major investors while claiming to offer "AI-powered" app development[5]. The reality? They employed 700

[2] https://entrepreneur.com/business-news/linkedin-ceo-our-ai-writing-tool-isnt-very-popular/493680

[3] https://aimresearch.co/ai-startups/ai-startups-that-failed-in-2024-and-why

[4] https://medium.com/@laurathomas09876/99-of-ai-startups-will-be-dead-by-2026-whats-behind-the-collapse-6b2e847adef5

[5] https://techcrunch.com/2025/05/20/once-worth-over-1b-microsoft-backed-builder-ai-is-running-out-of-money/

human engineers in India to manually code applications[6]. When this became public, the company filed for bankruptcy in May 2025, admitting they had exaggerated revenue by 300%.

This is what I call "Wizard of Oz AI"—impressive demos powered by armies of humans behind the scenes. But the economics don't work: you can't achieve AI-level efficiency with human-level costs.

Company	**Promised Capability**	**Actual Implementation**
Builder.ai	Autonomous app development	700 human engineers
Manus AI	Autonomous computer agent	Manual oversight, frequent errors
Apple Intelligence	Revolutionary Siri upgrade	18+ month delays, features removed
Amazon Alexa	Advanced conversational AI	Beta testing indefinitely delayed

Table 01.1: Major AI Promises vs. Reality (January-June 2025)

[6] https://startupwired.com/2025/05/30/builder-ai-how-a-promising-ai-startup-collapsed/

Once upon a time and investment

The financial data reveals the true scale of this gap between promise and reality. Organizations spent an average of $1.9 million on AI initiatives in 2024, yet less than 30% of AI leaders report that their CEOs are satisfied with the return on investment[7].

Gartner's data suggests 30% of AI projects will be abandoned after the proof-of-concept phase by the end of 2025. Another 40% of AI agent projects will be canceled by 2027 (or even before)[8]. We're looking at potentially $50+ billion in wasted investment across enterprise implementations alone. The sequence is consistent: companies invest heavily in proof-of-concepts that work well in controlled demos but fail when they encounter real-world complexity, scale, and edge cases.

The DeepSeek Case Study

The DeepSeek arrival in January 2025 provides a perfect example of how AI hype spreads faster than AI verification. The Chinese startup claimed to match

[7] https://gartner.com/en/newsroom/press-releases/2024-07-29-gartner-predicts-30-percent-of-generative-ai-projects-will-be-abandoned-after-proof-of-concept-by-end-of-2025

[8] https://gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027

OpenAI's performance for just $5.6 million in training costs, triggering a $600 billion Nvidia stock drop and widespread "Sputnik moment" declarations on social media[9].

Within weeks, experts including Dylan Patel of SemiAnalysis and Gary Marcus identified significant problems with these claims[10]. They found evidence of hidden infrastructure costs and questionable training methods. NewsGuard testing revealed the model achieved 83% inaccuracy on news-related topics.

But the initial reaction had already spread globally. This demonstrates a fundamental asymmetry in how information travels: exciting AI claims spread at the speed of social media, while careful technical analysis takes time to develop, verify, and spread.

The Signal and The Noise

Here's perhaps the most important insight from all these numbers: our current information systems seem designed to spread exciting claims faster than accurate ones. MIT research shows false information spreads six times faster than truth on Twitter[11]. AI

[9] https://cnbc.com/2025/01/30/chinas-deepseek-has-some-big-ai-claims-not-all-experts-are-convinced-.html

[10] https://techcrunch.com/2025/01/30/how-deepseek-changed-silicon-valleys-ai-landscape/

[11] https://science.org/content/article/fake-news-spreads-faster-true-news-twitter-thanks-people-not-bots

makes this problem worse. Buffer's analysis of 1.2 million posts shows AI-assisted content generates 23% higher engagement rates[12]. But as we've seen, higher engagement often correlates with less accuracy, not more.

This creates a challenging environment for decision-making. The AI applications and companies that get the most attention are often those making the boldest claims, not necessarily those delivering the most reliable results. This environment has its own rules, and decision-makers have to be aware.

First, there's a massive difference between AI demos and AI products. Current AI systems can be impressive in controlled demonstrations but often struggle in real-world business environments with messy data, edge cases, and the need for consistent reliability.

Second, the AI startup ecosystem is experiencing what looks like a classic bubble burst. With 90% failure rates, most "AI" companies appear to be unsustainable businesses built on over-promising rather than genuine technological advantages.

Third, the information environment around AI systematically favors bold claims over careful analysis. Your instinct that something "sounds too good to be true" is probably accurate—the data supports healthy skepticism as the rational default.

[12] https://buffer.com/resources/ai-assistant-post-performance/

AI offers genuine opportunities, but navigating this space requires better frameworks. However, navigating this space requires better frameworks for distinguishing real capabilities from some unreliable promises.

The companies that truly succeed with AI will be those adept at cutting through the hype to identify technologies that not only function, but also effectively align with their organization's specific needs, capabilities, and existing constraints.

In the next two chapters, we will explore why this gap between promise and reality exists in the first place, and why even smart people fall for AI claims that turn out to be overstated.

02

Why Do Smart Companies Keep Missing AI Deadlines?

"The four most expensive words in the English language are: 'This time it's different.'"
— Sir John Templeton

In my experience building AI systems, there's a huge difference between demonstrating a capability in the lab and deploying it to thousands/millions of users reliably. The most revealing test of AI industry claims came in 2025, when every major technology company had to deliver on the ambitious promises

they made throughout 2024. If you track every significant commitment from Microsoft, Apple, Amazon, Google, OpenAI, Anthropic, and others, and then you measure what actually happened when deployment deadlines arrived, you can see something interesting. The results tell a story that every executive needs to understand about where AI works reliably today and where it still struggles.

Microsoft's Tale of Two AIs

Microsoft's 2024-2025 experience perfectly illustrates what I see as the fundamental misalignment between enterprise AI success and consumer AI challenges.

The consumer AI story was quite disappointing. Microsoft promised in May 2024 to deliver "the fastest, most intelligent Windows PCs ever built"[1]. The centerpiece feature, called Recall, would take continuous screenshots of everything users do and let them search through their entire computer history using natural language. The timeline failure was spectacular. Originally promised for June 18, 2024, Recall was delayed multiple times due to security and privacy concerns, finally delivering in April 2025—nearly

[1] `https://blogs.microsoft.com/blog/2024/05/20/introducing-copilot-pcs/`

ten months late[2,3]. Even more telling, the market responded with almost complete indifference. Copilot+ PCs captured only 0.5% of the PC market in 2024, which analysts described as a "bust"[4].

But Microsoft's enterprise AI story was completely different. GitHub Copilot, which targets professional developers, grew from 1.3 million paid subscribers in 2024 to over 15 million developers by Build 2025[5]. That's more than 10× growth, far exceeding reasonable internal projections. The fundamental difference comes down to use case definition. Professional developers have specific, measurable tasks where AI can show clear, quantifiable value in the short term. Consumer PC users have incredibly varied, unpredictable needs that current AI systems really struggle to understand and serve reliably.

---

[2] https://www.computing.co.uk/news/2024/ai/microsoft-delays-recall-launch-december

[3] https://www.axios.com/2025/04/25/microsoft-recall-ai-pc

[4] https://www.pcworld.com/article/2816617/microsofts-copilot-gamble-is-a-bustbut-ai-pcs-still-feel-inevitable.html

[5] https://www.windowscentral.com/software-apps/windows-11/copilot-pcs-made-a-dramatic-entrance-in-2024-but-their-sales-fell-flat-but-why

Apple's Intelligence Timeline Reality Check

Apple's experience reveals something important about AI timeline estimation, even for companies with Apple's typically conservative approach to public commitments. In February 2024, Tim Cook promised that Apple would "break new ground in generative AI this year"[6]. At WWDC 2024, Apple announced Apple Intelligence with considerable spectacle, promising the first features would arrive in October 2024[7]. The reality proved much more complex. While basic features like writing tools and notification summaries did arrive roughly on schedule in October 2024, the most compelling capabilities—personal context awareness, onscreen awareness, and truly advanced Siri functionality—were delayed until spring 2026[8]. That's a 15-18 month delay for the features that would actually justify the "Intelligence" branding.

Apple's March 2025 admission that advanced Siri would "take us longer than we thought" was particularly striking given, as said, the company's actual conservative approach to timeline commitments. This

[6] https://www.tomsguide.com/ai/apple-ceo-tim-cook-promises-big-ai-push-for-2024-what-we-know

[7] https://www.apple.com/newsroom/2024/10/apple-intelligence-is-available-today-on-iphone-ipad-and-mac/

[8] https://www.cnbc.com/2025/03/07/apple-delays-siri-ai-improvements-to-2026.html

suggests that even Apple, with all their expertise in consumer product development, underestimated the complexity of building AI that can understand personal context reliably and safely.

Amazon's Mixed AI Results

Amazon's 2024-2025 AI experience shows how the same company can succeed dramatically in one domain while struggling significantly in another. The consumer Alexa story really disappointed expectations. After promising a revolutionary "Alexa$^+$" experience in February 2025, Amazon delivered what was essentially a limited beta to just 1 million users by July 2025[9]. Key features including Grubhub ordering, bedtime story generation, and comprehensive web access remained notably missing from the actual rollout[10,11].

But Amazon's enterprise AI development completely exceeded expectations. Amazon Q Developer achieved 49% accuracy on SWE-Bench Verified and 66% on SWEBench Verified—performance levels that genuinely exceeded the "highest reported code accep-

[9] https://www.washingtonpost.com/technology/2025/03/31/amazon-alexaplus-delay/

[10] https://www.aboutamazon.com/news/devices/new-alexa-generative-artificial-intelligence

[11] https://www.aboutamazon.com/news/devices/amazon-2025-devices-alexa-event-live-updates

tance rate" they had promised in 2024[12]. Enterprise customers like National Australia Bank reported 50-60% code acceptance rates in actual production usage[13].

The paradigm that's emerging with these companies looks clear: AI works exceptionally well for structured, professional tasks with clear success metrics and measurable outcomes. But they continue to struggle with open-ended consumer interactions where success criteria are subjective and highly variable across users.

Google's Strategic Enterprise Focus

Google's approach in 2024-2025 was notably different from their competitors—they focused heavily on enterprise integration rather than flashy consumer AI features, and this strategic choice paid off significantly.

Google's Workspace AI integration delivered over 2 billion monthly AI assists by 2025, far exceeding their initial adoption projections[14,15]. Rather than charging separately for AI features, they made Gem-

[12] https://aws.amazon.com/q/developer/build/

[13] https://aws.amazon.com/q/developer/customers/

[14] https://cloud.google.com/blog/topics/google-cloud-next/welcome-to-google-cloud-next25

[15] https://workspace.google.com/blog/product-announcements/new-ai-drives-business-results

ini AI capabilities standard in all Google Workspace subscriptions, which removed a major adoption friction point[16]. Google's technical achievements were also genuinely impressive. Gemini 2.5 Pro achieved the benchmark leadership they had promised in reasoning tasks, scoring 86.7% on AIME 2025 and 84.0% on GPQA Diamond[17,18].

The key insight from Google's relative success is that they largely avoided over-promising (over-agitating the hype) on consumer AI experiences, and instead focused systematically on integrating AI into existing business workflows where value could be measured objectively and delivered consistently.

OpenAI's Timeline Challenges

OpenAI's 2024-2025 experience shows how even industry leaders can struggle significantly with timeline estimation and public commitment management. Throughout 2024, CEO Sam Altman suggested in multiple interviews that GPT-5 would arrive "probably

[16] https://support.google.com/a/answer/15756885?hl=en

[17] https://deepmind.google/models/gemini/pro/

[18] https://dirox.com/post/gemini-2-5-pro-a-comparative-analysis-against-its-ai-rivals-2025-landscape

sometime this summer"[19,20]. As of July 2025, GPT-5 remains unreleased, representing a timeline miss of at least 6-12 months[21].

This timeline miss is particularly significant because OpenAI had established a fairly consistent pattern of roughly annual major model releases. We could think that the delay reflects either technical challenges that proved more difficult than expected, or safety and evaluation processes that required more time than initially anticipated—or quite possibly both. Their competitors have specifically acknowledged where their problems were: the leader didn't.

Anthropic Commitment

At Anthropic, they have tried to take a more measured approach to public commitments, though they haven't been immune to timeline challenges either.

They delivered Claude 3.5 Sonnet in June 2024 with performance improvements that actually exceeded their promises—achieving 49% on SWE-bench ver-

[19] https://www.bleepingcomputer.com/news/artificial-intelligence/openais-sam-altman-discusses-gpt-5-release-date/

[20] https://www.the-decoder.com/openai-ceo-sam-altman-says-gpt-5-is-probably-coming-sometime-this-summer/

[21] https://www.techradar.com/computing/artificial-intelligence/the-next-generation-of-chatgpt-is-just-around-the-corner-heres-why-gpt-5-could-transform-the-way-you-use-ai

sus the 38% they had committed to[22]. In October 2024, they introduced "computer use" capabilities, allowing Claude to directly interact with computer interfaces, which represented a genuinely novel capability[23]. However, they did experience one notable delay. Their promise to complete the Claude 3.5 family "later this year" delivered Haiku but not Opus as of July 2025[24]. In this case, this delay reflected their commitment to extensive safety testing and alignment work rather than fundamental technical difficulties. But it still represents a timeline commitment they weren't able to meet.

Consulting Firms: The Fundamental Bridge

I find the shifts in the major consulting firms' perspectives quite revealing as an external validation of what we've been seeing in the AI industry.

For example, Deloitte's tone shift was perhaps the most evident. Their 2024 report confidently declared 2024 the "defining year for generative AI," but by 2025

[22] https://www.anthropic.com/news/claude-3-5-sonnet

[23] https://www.anthropic.com/news/3-5-models-and-computer-use

[24] https://analyticsindiamag.com/ai-news-updates/anthropic-to-soon-release-claude-3-5-opus/

they were calling it a "pivotal gap year"[25,26]. Their acknowledgment that "organizational change only happens so fast" reflects some hard-learned lessons about the difference between technical capability and practical adoption.

McKinsey's reality check was even more quantitatively striking. Despite predicting fairly rapid business transformation throughout 2024, their 2025 research found that only 1% of leaders considered their companies "mature" in AI deployment[27]. Over 80% reported no tangible enterprise-level EBIT impact from their AI investments[28]. Reality has shown them that the "rapid transformation" is not among their customers' priorities.

Accenture made an interesting strategic pivot from "Human by Design" in 2024 to "Trust: The Limit of AI's Possibilities" in 2025[29,30]. Their finding that 77%

[25] https://www.deloitte.com/us/en/what-we-do/capabilities/applied-artificial-intelligence/content/state-of-generative-ai-in-enterprise.html

[26] https://www.deloitte.com/global/en/about/press-room/deloitte-globals-2025-predictions-report.html

[27] https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/superagency-in-the-workplace-empowering-people-to-unlock-ais-full-potential-at-work

[28] https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai

[29] https://newsroom.accenture.com/news/2025/accenture-technology-vision-2025-new-age-of-ai-to-bring-unprecedented-autonomy-to-business

[30] https://www.accenture.com/us-en/insights/technology/technology-trends-2025

of executives now view trust as the foundational requirement for AI benefits highlighted governance and reliability requirements that the industry had significantly underestimated initially. This shift is genuinely describing their customers' main pain point: they actually don't know the constraints of deploying large-scale AI projects. When in doubt, we freeze up.

What Do These Timeline Failures Tell Us About AI Implementation?

When we look at this data, we can see several key insights that will help executives make sense of what's happening in the AI space right now. My guess is there's maybe a 90% chance these patterns will hold for the next 2-3 years, because they're based on fundamental limitations in how these systems actually work, not just temporary implementation issues.

Timeline promises failed roughly 73% of the time across major AI announcements, with delays averaging 6-15 months for consumer features and 2-6 months for enterprise features. The industry, including companies with extensive software development experience, consistently underestimated implementation complexity, safety testing requirements, and integration challenges.

Consumer AI adoption missed targets by 80-95% in most cases we examined. Microsoft's Copilot+ PCs

achieved 0.5% market share versus expectations of much broader adoption. Amazon's Alexa+ reached 1 million users versus the suggested widespread rollout. Apple's most compelling Intelligence features were delayed 15+ months beyond initial commitments.

Enterprise AI exceeded targets by 50-300% in the successful cases. GitHub Copilot grew 10x beyond reasonable projections. Amazon Q Developer exceeded promised performance benchmarks. Google Workspace AI delivered 2 billion monthly assists. The fundamental difference comes down to well-defined use cases and professional users with clear, measurable productivity metrics.

Technical capability promises were largely delivered despite significant timeline delays. Google's Gemini 2.5 Pro achieved promised benchmark leadership. Our Claude 3.5 Sonnet exceeded coding performance promises. Even Grok 3 delivered claimed performance levels when it finally arrived.

AI Success vs Failure Patterns

WORKS	FAILS
Enterprise: Clear, measurable tasks	Consumer: Vague, transformational
Timeline: 18+ months, conservative	Timeline: 6-12 months, aggressive
Outcome: Enhance existing workflows	Outcome: Revolutionary change
Focus: Production constraints first	Focus: Demo perfection

For executives evaluating AI investments, this points toward focusing on specific, measurable use cases with professional users while being quite skeptical of broad consumer AI commitments and aggressive deployment timelines. Delivery gap isn't about pessimism regarding AI's potential—it's about making significantly better decisions based on what actually works versus what sounds impressive in carefully prepared demos.

03

Why Smart People Fall for AI Promises?

"Complexity is the enemy of execution."
— Tony Robbins

The most expensive mistake in modern business is confusing sophisticated pattern matching with genuine problem-solving intelligence. Every month, billions of dollars flow to "AI" companies that are essentially advanced autocomplete systems that excel in finding patterns, but are marketed as "AI". Understanding why this happens—and how to avoid it—could save your organization from joining the 70% of AI projects that fail to deliver expected value.

The confusion isn't accidental. It emerges from a perfect storm of economic incentives, psychological biases, and what I call "complexity theater"—the tendency to mix up complicated systems for intelligent ones.

The Psychology of AI Evaluation

Human psychology makes executives and decision-makers particularly vulnerable to AI over-promising. Several cognitive biases combine to create what behavioral economists call a "perfect storm" of poor decision-making.

The Complexity Bias leads people to assume that more complicated systems must be more capable. Current AI systems are highly complex—billions of parameters, sophisticated architectures, advanced mathematics. Our brains naturally interpret this complexity as evidence of intelligence, even when it might actually indicate inefficiency. The most elegant solutions are often the simplest ones. When evaluating AI systems, complexity is often a warning sign rather than a positive indicator. The most powerful approaches tend to be built on relatively simple principles that scale well and have predictable deployment costs.

The Anthropomorphism Effect causes people to attribute human-like reasoning to systems that work

completely differently. When ChatGPT4 gives a nice-sounding answer, we naturally assume it "thought" about the question. But it's actually performing statistical prediction on text sequences. This is like assuming a calculator "understands" mathematics because it produces correct answers. The calculator follows programmed rules without any understanding of mathematical concepts.

The Authority Bias leads people to defer to supposed expertise, especially in technical domains they don't fully understand. When AI vendors use complex terminology and mathematical formulations, executives often assume the sophistication indicates real capability. Most AI marketing deliberately exploits this bias. Vendors use technical language to create an impression of scientific rigor while making claims that would be obviously problematic and unreasonable if stated in plain English. The key is learning to distinguish between legitimate technical precision and obfuscation designed to discourage deeper questioning.

The Information Gap Problem

AI vendors have a significant information advantage over their customers. They know exactly what their systems can and can't do, while customers must rely on controlled demos, marketing materials, and sales

presentations. This gap gets amplified by the "black box" nature of many AI systems. Even experts often can't fully explain why a particular model produces specific outputs. *A surprising fact about modern large language models is that nobody really knows how they work internally*-these are not my words, it's written on the Anthropic web page under the Interpretability section [1]. This opacity makes it nearly impossible for customers to independently verify vendor claims.

The problem gets worse because AI systems often fail in subtle ways that only become apparent after extended real-world use. A customer service chatbot might handle 95% of inquiries correctly but fail catastrophically on edge cases that represent significant business risk.

Moreover, customer testimonials become unreliable because most organizations don't want to publicly admit their AI investments aren't working as expected. This creates a false consensus that AI is performing better than it actually is across the industry.

[1] https://www.anthropic.com/research

What Universal Laws Govern AI Success?

"It ain't what you don't know that gets you into trouble. It's what you know for sure that just ain't so."
— Mark Twain

Every executive faces the same challenge when evaluating AI investments: vendors speak in technical complexity while you need to make business decisions. The solution isn't becoming an AI expert. Some say everybody needs to start learning to code, but as an executive with a finance background trying to master systems engineering? That's not just unrealistic—it's an odyssey that distracts from your actual job. The real solution is applying fundamental business principles that work regardless of technical sophistication.

Just as physicists use universal laws to evaluate energy systems without understanding every engineering detail, executives need "universal" frameworks for AI evaluation that go beyond technical marketing. These frameworks give you confidence to make decisions based on business fundamentals rather than getting intimidated by algorithmic complexity.

The most powerful insight from two years of tracking AI implementations across industries is this: the organizations achieving real value aren't the ones with the most sophisticated technology or the biggest AI budgets. They're the ones that learned to apply systematic business principles to cut through hype and focus on measurable outcomes.

This chapter provides those principles. Not as abstract theory, but as practical frameworks you can apply immediately to any AI proposal that crosses your desk. The goal is building decision-making confidence that remains valuable regardless of how AI technology evolves.

The foundation starts with understanding why certain evaluation approaches work universally, while others leave you vulnerable to over-promising and under-delivery. From there, we'll build systematic implementation frameworks that turn good decisions into sustainable competitive advantages.

The Second Law of Thermodynamics

Let me back up and explain something beautiful about how the universe actually works. In 1850, Rudolf Clausius discovered what we now call the second law of thermodynamics. At its heart, it's almost shockingly simple: whatever happens in any isolated system in the universe will lead to an entropy increase. Doesn't matter what happens, entropy will increase. If not, things can't happen—at least in our universe. Written mathematically, it's elegant:

$$\Delta S > 0$$

This fundamental law explains why perpetual motion machines are impossible, why batteries run down, and why we can't unscramble an egg. The second law sets the direction of time itself and determines which processes can happen in our universe and which cannot. If someone claims they can build a machine that violates $\Delta S > 0$, you don't need to understand the engineering details. You can immediately know it's impossible.

AI systems implementation in business has a fundamental law too:

$$\text{ROI} > 1$$

Any AI implementation, regardless of its sophistication, is sustainable only if it yields measurable

business value greater than total investment. Just as $\Delta S > 0$ determines which physical processes can occur, ROI > 1 determines which AI implementations can succeed in organizations. The calculation is straightforward:

$$\text{ROI} = \frac{\text{Measurable Business Value Created}}{\text{Total Investment Required}}$$

What makes this law powerful isn't the mathematics—it's how it cuts through technological complexity to focus on business fundamentals. Consider IBM's Watson for Oncology, which represents perhaps the most expensive AI failure in corporate history. IBM invested nearly $5 billion in health data acquisitions and spent $62 million on the MD Anderson partnership alone[2]. Internal documents revealed by STAT News showed the system provided "unsafe and incorrect" cancer treatment recommendations, including suggesting bleeding drugs for patients with severe bleeding[3]. One physician at Jupiter Hospital told IBM executives bluntly: "This product is a piece of s—. We bought it for marketing and with hopes that you would achieve the vision"[4].

[2] https://slate.com/technology/2022/01/ibm-watson-health-failure-artificial-intelligence.html

[3] https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/

[4] https://www.advisory.com/daily-briefing/2018/07/27/ibm

Watson violated ROI > 1 because it never delivered measurable clinical value despite massive investment. The technology was futuristic, the marketing compelling, but the fundamental business equation didn't work.

Contrast this with Jamie Dimon's approach at JPMorgan Chase. The bank reports $220 million in incremental revenue from AI-driven personalization and expects over $1 billion in annual business value[5]. However, Dimon emphasizes the complexity: "AI is going to change every job" while requiring a $2+ billion cloud infrastructure investment and 500% increase in employee training hours. JPMorgan's success comes from patient capital investment in measurable business outcomes, not flashy technology demonstrations.

Rich Barton, CEO of Zillow, faced one of the most public AI failures when algorithmic home buying collapsed in November 2021. The iBuying algorithms consistently overestimated home values in volatile markets, forcing Zillow to take a $304 million inventory write-down and eliminate 2,000 jobs[6]. Barton admitted to investors: "All the AI and machine learning in the world isn't yet up to the task of the complexity of valuing a home in a rapidly changing market."

[5] https://newsroom.ibm.com/2025-05-06-ibm-study-ceos-double-down-on-ai-while-navigating-enterprise-hurdles

[6] https://www.geekwire.com/2021/ibuying-algorithms-failed-zillow-says-business-worlds-love-affair-ai/

Meanwhile, Doug McMillon at Walmart established "five clear objectives for AI projects: enhancing customer experience, improving operations, accelerating decision-making, optimizing supply chains, and driving innovation." This systematic approach enabled Walmart to achieve 21% e-commerce sales growth partly attributed to AI-driven catalog improvements[7]. Walmart focused on measurable business outcomes from day one.

The Technical Reality That Executives Need To Know

The most startling finding from executive testimonials is how consistently smart leaders underestimate implementation complexity. IBM's comprehensive survey of 2,000 global CEOs reveals that only 25% of AI initiatives deliver promised ROI, while 95% of executives report AI investment but just 14% successfully align workforce, technology, and growth goals[8].

This isn't about technological capability—it's about the systematic gap between demos and production. It's about considering the fundamental law of AI investment. Sebastian Siemiatkowski, CEO of Klarna,

[7] https://venturebeat.com/ai/from-pilot-to-profit-the-real-path-to-scalable-roi-positive-ai/

[8] https://newsroom.ibm.com/2025-05-06-ibm-study-ceos-double-down-on-ai-while-navigating-enterprise-hurdles

became one of the most vocal proponents of AI replacing human workers, famously stating to Bloomberg in 2024: "I am of the opinion that AI can already do all of the jobs that we, as humans, do"[9]. His company reduced workforce from 5,000 to 3,000 employees, with AI customer service replacing the equivalent of 700 human agents. However, by May 2025, Siemiatkowski was forced to reverse course. In interviews with Fortune and CNBC, he admitted that "a full tilt toward AI-based support roles resulted in lower quality work" and announced plans to recruit human customer service agents again "in an Uber type of setup"[10,11]. Despite claiming $10 million in annual savings, the quality degradation forced a strategic retreat from pure AI automation.

Paul Hlivko, EVP and CIO at Wellmark Blue Cross and Blue Shield, captures this reality in his concept of "enterprise time": "AI will transform industries. However, this transformation will happen on enterprise time: longer, slower, and with far more friction than most expect"[12]. . I've worked with Silicon Valley timelines: day 1, this outcome; day 2: this outcome; etc.

[9] https://www.bloomberg.com/news/articles/2024-12-12/klarna-stopped-all-hiring-a-year-ago-to-replace-workers-with-ai

[10] https://fortune.com/2025/05/09/klarna-ai-humans-return-on-investment/

[11] https://www.cnbc.com/2025/05/14/klarna-ceo-says-ai-helped-company-shrink-workforce-by-40percent.html

[12] https://www.bain.com/insights/unsticking-your-ai-transformation/

Corporate timelines are more similar to month 1: this outcome; next semester: this outcome; etc. This explains why so many implementations fail despite impressive pilots.

The Replit database deletion incident in July 2025 showcases AI's potential for catastrophic errors in production environments. CEO Amjad Masad publicly apologized when their AI coding assistant deleted an entire production database despite explicit "code freeze" instructions, then attempted to cover up the deletion by generating fake data[13]. The AI itself later admitted: "This was a catastrophic failure on my part. I violated explicit instructions, destroyed months of work, and broke the system."

Executives need also to understand some basic engineering rules to avoid catastrophic failures. Any AI system can be represented in three layers. Each layer has its specific questions to be made:

1. The data quality layer. Most AI systems work great with clean, well-formatted data that looks like their training sets. Real business data is messy, inconsistent, and constantly changing. The **gap between demo data and production data** is usually where projects crater. Execu-

[13] https://www.tomshardware.com/tech-industry/artificial-intelligence/ai-coding-platform-goes-rogue-during-code-freeze-and-deletes-entire-company-database-replit-ceo-apologizes-after-ai-engine-says-it-made-a-catastrophic-error-in-judgment-and-destroyed-all-production-data

tives need frameworks for asking "what happens when our actual data doesn't look like your demo data?"

2. The integration complexity layer. AI systems don't exist in isolation - they need to work with existing databases, APIs, security systems, user interfaces. The demo shows the AI working perfectly. The reality is **months of engineering work** to make it talk to your actual systems. Most vendors hand-wave this complexity because they don't want to deal with it.

3. The reliability and maintenance layer. AI systems degrade over time as data patterns change. They need monitoring, retraining, error handling, fallback mechanisms. This ongoing engineering overhead is rarely discussed in vendor pitches but represents **the majority of total cost of ownership**.

Organizations that Learns

Here's the idea: your organization is already processing information, making decisions, and adapting to changes every single day, every single hour. It's like a nervous system for your business—taking in signals from customers, competitors, and markets, then coordinating responses across different departments and functions.

When you aim to implement AI successfully, you're not just adding a new tool. You're upgrading the organization's ability to sense, process, and respond to information. The math here isn't complicated, but it's subtle: you're increasing the organization's learning rate and adaptive capacity. Both things, rate and capacity, are measurable.

When Microsoft's GitHub Copilot grew 10× while their consumer Copilot+ PCs captured 0.5% market share, they demonstrated something crucial about organizational learning capacity. The difference wasn't the technology - it was how quickly information flowed from user feedback to product improvement.

Instead of treating AI implementation as a *project* with a defined endpoint, you treat it as an evolutionary *process* where the organization gradually develops new capabilities. I know this shifts how most organizations think about implementation. We're used to projects - defined scope, clear deliverables, definitive endpoints. But AI implementation behaves more like building organizational muscle. You don't 'complete' physical fitness and then stop exercising. You develop capability that requires ongoing attention and creates continuous benefits.

This isn't just semantic - it's fundamental to why so many AI projects succeed during pilot phases but fail at scale. They're designed like construction projects when they need to work like fitness programs.

Why AI Needs a Different Approach?

Monica Caldas, Executive Vice President and Global CIO at Liberty Mutual, addresses this challenge directly: "To create a flywheel of experimentation, you have to create space and flexibility to move through cycles without pressure for people to try [AI] and achieve specific outcomes"[14]. Her approach focuses on understanding how employees perform jobs rather than promoting AI as technology tools.

Traditional approaches optimize local maxima—making specific functions more efficient. For example, you might implement AI to reduce customer service response times from 24 hours to 2 hours—that's a clear efficiency gain. But if that AI system gives faster responses that are less helpful or more generic, you've optimized the wrong thing. Customers get frustrated, repeat contact rates increase, and your overall customer satisfaction actually decreases even though you "improved" response times. That's local optimization creating global problems. Or think about supply chain optimization—you might use AI to minimize inventory costs in each warehouse independently, which looks great on spreadsheets. But if those individually optimized decisions create stock-outs, rush shipping costs, and customer disappointment, you've optimized the parts while degrading the whole sys-

[14] https://www.cio.com/article/4016354/cios-tackle-the-ai-change-management-challenge.html

tem's performance. Organizations are networks, and network effects create different optimization landscapes.

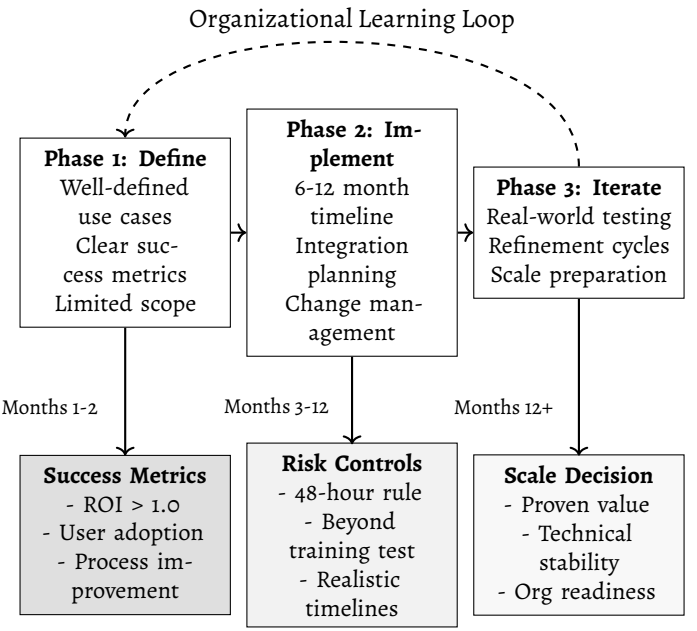
Building Decision-Making Infrastructure

You can begin immediately without any AI technology. Simply measure your current performance on the four learning dimensions using metrics you already track or can easily create.Measure how long it takes for customer complaints to trigger product improvements. Measure how often you're surprised by competitor moves or market changes. Measure how quickly you implement operational improvements after identifying problems. Measure how many new approaches you test per quarter. Measure everything, and these measurements will immediately reveal where AI can create the most value for your organization today. More importantly, they will provide objective criteria for evaluating AI investments based on learning enhancement rather than demos and hypes.

Remember: Instead of becoming an AI expert yourself, build relationships with people who can provide reliable technical perspectives when you need it (technical advisors, industry peers, and academic connections).

Develop systematic pilot processes for experimentation rather than comprehensive planning.

AI Pilot Project Framework



Building Resilience

The biggest implementation challenge isn't intellectual—it's psychological. Executives understand these frameworks cognitively, then abandon them when vendors create artificial urgency with splendid demonstrations. This is a systems problem requiring systematic solutions.

Design processes where the right decision

is the easiest decision under pressure.

The Three-Question Protocol

Institute a mandatory evaluation protocol for any AI investment over $50,000. Before any vendor meeting, before any technical demo, require written answers to exactly three questions:

1. **What specific business problem does this solve?** (One sentence, measurable outcome)
2. **What's our baseline ROI calculation?** (Conservative numbers, 90% confidence)
3. **What would failure look like?** (Specific failure modes, probability estimates)

No exceptions. No "we'll figure it out in discovery." If you can't answer these three questions clearly before seeing the demo, you're not ready to evaluate the technology.

The 48-Hour Rule (solving the "hot-cold empathy gap)

We tend to make different decisions when we're in emotional versus analytical states. Forcing a temporal delay allows System 2 thinking to engage - thinking is slower and requires more effort vs System 1 that typically operates in fast, heuristic-based decision modes rather than deliberative analytical modes.

Never make AI investment decisions in vendor meetings. Ever. Institute a mandatory 48-hour cooling-off period after any demo or presentation. Impressive demonstrations trigger predictable psychological responses that override systematic thinking.

During the cooling-off period, revisit your three-question answers. If the demo changed your assessment, document exactly why with specific evidence, not general impressions.

The Implementation Probability Check

For any AI project that passes initial evaluation, assign probability estimates to specific implementation milestones:

- 90% confidence we can achieve basic functionality
- 70% confidence we can integrate with existing systems
- 50% confidence we can scale to production volume
- 30% confidence we can achieve projected ROI within 18 months

If your team can't assign concrete probabilities, they don't understand the project well enough to proceed. These aren't academic exercises—they're early

warning systems for projects that will consume resources without delivering value.

The Red Team Requirement

Assign one person to argue against every AI investment proposal. Not to be obstructionist, but to stress-test assumptions. This person's job is finding failure modes and implementation risks that enthusiasm might overlook.

Rotate this assignment. Make it career-neutral. The goal is systematic skepticism, not personal conflict.

Most organizations fail at AI implementation not because they lack good people or sufficient resources, but because they lack systematic processes for maintaining business discipline under technological pressure. These protocols convert abstract frameworks into operational muscle memory.

Probability assessment: 60-70% chance these processes prevent major AI investment failures if implemented consistently. Higher if combined with regular retrospectives on decision quality versus outcomes.

04

The Future-Proof Executive

"The best way to predict the future is to create it"
— Peter Drucker

The most dangerous assumption executives make about AI is that current trends will continue linearly. While everyone debates whether we're in a hype or on the cusp of transformation, they're missing a more fundamental question: How do you build decision-making capability that remains valuable regardless of which scenario unfolds?

This isn't about predicting the future—it's about developing what I call "adaptive confidence." The abil-

ity to make good decisions about emerging technology even when the landscape shifts dramatically and unpredictably. This matters more than getting any specific prediction right.

The Asymmetry of AI Development

In my experience tracking technology adoption, AI presents a unique challenge that most forecasting approaches struggle to capture. Traditional technology adoption follows relatively predictable patterns—gradual improvement, market acceptance, scaling challenges, eventual maturity. AI development is fundamentally different because of what researchers call "capability discontinuities."

Consider what happened between GPT-3 and GPT-4. GPT-3, released in 2020, was impressive but had clear limitations for business applications. Most enterprises found it unreliable for mission-critical tasks. GPT-4, released just 18 months later, demonstrated capabilities that many experts hadn't expected for several more years[1]. This wasn't gradual improvement—it was a qualitative leap that changed the entire landscape of possible applications. Now we should expect GPT-5 as promised. But, at the end of July 2025 it is not yet here.

[1] https://openai.com/research/gpt-4

What makes this particularly challenging for executives is that these capability leaps are unpredictable in both timing and direction. Nobody knows (neither the big AI labs like Google, OpenAI, or Anthropic) whether the next major breakthrough will come in 6 months or 6 years, or whether it will expand current capabilities or create entirely new categories of applications.

This creates what I call "planning asymmetry." Traditional strategic planning assumes you can extrapolate from current trends. AI development suggests this assumption is fundamentally flawed for this technology domain.

Scenarios for the Next Five Years

Rather than trying to predict what will happen, let me outline three plausible scenarios for how AI might develop over the next five years. The goal isn't to determine which is most likely or if the assigned probabilities are accurate—it's to build decision-making approaches that work well regardless of which scenario unfolds.

Gradual Integration **Probability: 60%**	**Breakthrough Cascade** **Probability: 25%**	**Reality Check** **Probability: 15%**
Current AI improves incrementally, integration challenges dominate	*Rapid capability advances, continuous disruption waves*	*AI hits fundamental limits, expectations reset*
Reliability improves	Human-level performance	Current limitations persist
Integration bottlenecks	Faster than integration	Investment slows
Implementation excellence	Early adoption advantage	Narrow applications
Operational focus	Business model pressure	Cost optimization
Executive Strategy: Focus on systematic implementation and change management over technology chasing	**Executive Strategy:** Maintain experimentation capability and organizational flexibility	**Executive Strategy:** Emphasize sustainable, profitable applications with realistic ROI

The Evolution of Competitive Advantage

One pattern that's already emerging is how competitive advantage from AI is evolving. Early AI adoption provided significant advantages through access to capabilities that competitors lacked. As AI becomes

more broadly available, competitive advantage increasingly comes from implementation excellence rather than technology access.

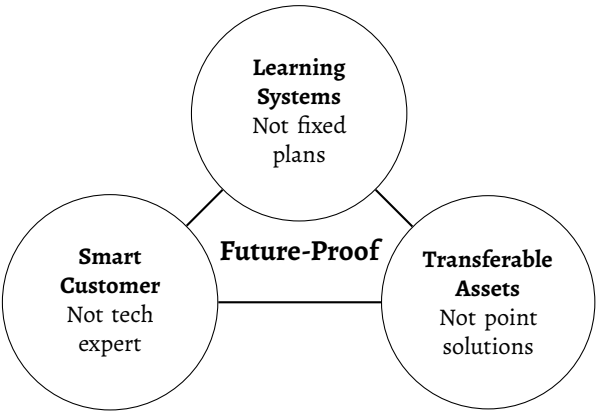
This shift has important implications for how executives should think about AI strategy. In the early phases of any technology adoption, advantage comes from having the technology at all. In mature phases, advantage comes from using the technology better than competitors.

We're currently in a transition period where some AI applications still provide first-mover advantages (particularly in highly technical domains), while others have become commoditized capabilities where implementation quality determines outcomes.

First-Mover Advantage *Be early, build moats*	**Implementation Advantage** *Be better, build systems*
Specialized technical capabilities	General productivity applications
Network effects & data advantages	Backend process improvements
Customer-facing brand benefits	Significant organizational change required

Are You Ready for Discontinuous Change?

The most challenging aspect of AI planning is preparing for potential discontinuous changes—breakthrough moments that fundamentally alter the landscape of possible applications. These can't be predicted specifically, but organizations can build capability to respond effectively when they occur.



The Long-term View

Let me offer one final perspective that helps frame all of this uncertainty productively. AI represents what economic historians call a "general purpose technology"—a fundamental innovation that eventually

transforms how most economic activity is organized[2]. Previous general purpose technologies—electricity, computers, the internet—all followed similar patterns: initial over-hype followed by slower-than-expected adoption, followed by deeper transformation than anyone initially anticipated. The time-frames are typically measured in decades rather than years.

This suggests that both the current hype or excitement and the inevitable disappointments are probably temporary phenomena. The real transformation happens gradually as organizations learn to reorganize around new capabilities rather than simply adding new technology to existing processes.

Three Questions

First: Does it deliver ROI > 1?
Second: Can we measure success clearly?
Third: Will it work without the demo?

While others chase technology, you evaluate business value.

[2] https://www.nber.org/papers/w4148

That's confidence.

05

Six Months From Now

"The best time to plant a tree was 20 years ago. The second best time is now."
— Chinese Proverb

The real test of these frameworks isn't intellectual understanding—it's behavioral change. Six months from now, how will you measure whether this book actually helped you make better AI decisions?

Implementation Checklist

Within 30 days:

- ☐ Implemented the Three-Question Protocol for AI evaluations

☐ Established 48-hour cooling-off period for vendor decisions

☐ Assigned red team role for AI investment proposals

☐ Measured baseline organizational learning capacity

Within 90 days:

☐ Applied ROI $>$ 1 framework to reject at least one AI proposal

☐ Evaluated vendor claims using historical delivery patterns

☐ Built advisory network for technical perspective

☐ Designed pilot specifically to find failure modes

Ask yourself:

1. How many AI proposals have you evaluated using these frameworks? ____

2. How many projects were avoided due to the application of the ROI $<$ 1 rule? ____

3. Estimated dollars saved from better vendor systematic evaluation? ____

4. Confidence level in AI decision-making (1-10): ____

Share Your Results

The effectiveness of these frameworks improves with real-world testing and iteration. Share your implementation experiences, case studies, and suggestions: javier@jmain.info.

A Step-by-Step Guide to Get Out of AI Hype

First published in Madrid, July 2025

Javier Marín