

COMPLETE SERIES

# Return on Intelligence

AI and the Programmable Firm

Rebecca Powell

A nine-part essay on artificial intelligence, the bubble it is building, and the economy that will remain after it breaks.

May 2026

# Contents

**PREFACE**     **The Return**

The thesis, the reading contract, and the journey through the full eight-part argument.

---

**PART 1**     **Echoes**

Dotcom memory, paradigm shifts, and why real technologies produce the most dangerous bubbles.

---

**PART 2**     **Prototypes**

Early companies and interfaces are prototypes of futures they may not own. A prototype can be directionally right and economically doomed.

---

**PART 3**     **Moats**

Intelligence alone is not the moat. Durable moats may be context, distribution, workflow, identity, permissions, governance, execution rights, data and systems of record.

---

**PART 4**     **SaaS Was a Compromise**

SaaS won because custom software was too expensive. AI changes that cost curve, and the compromise gets repriced.

---

**PART 5**     **Firmware**

Firm-ware is the operating logic embedded into the company itself. AI makes that logic programmable.

---

**PART 6**     **Breaking Point**

The market does not stop believing in AI. It starts doubting who owns it.

---

**PART 7**     **After the Crash**

The crash is a repricing event. The mature AI economy begins after the first ownership story fails.

---

**PART 8**      **The New Power Map**

A layered view of the durable winners, pressured groups, and control points in the mature AI economy.

## PREFACE

# The Return

---

## Preface: The Return

This series begins with a claim that sounds contradictory until you sit with it for a while:

*The AI bubble will not burst because AI is fake. It will burst because AI is real.*

I do not mean the technology is weak. I mean it may be strong enough to redraw the map faster than the market can price it correctly.

That is what real paradigm shifts do. They make the future feel obvious before the ownership layer is settled. Capital rushes toward the most visible winners, but the mature economy often forms somewhere else.

That is why this series starts with dotcom memory rather than a valuation chart. I care less about whether AI is "overhyped" in the abstract than about whether the market is confusing the gateway with the destination, the early winners with the durable winners, and the first story with the final structure.

This is the argument running through every chapter:

- AI utility is already real.
- The first ownership story is probably not the final one.
- A crash, if it comes, may be a repricing, not a negation.
- The real AI economy may emerge after the obvious trade breaks.

Stanford's [AI Index](#) captures how economically and socially present AI already is, even while the long-term power structure remains unsettled.

## SERIES JOURNEY

## Return on Intelligence

AI and the Programmable Firm — read this sequence in order. Each chapter advances the same argument from a different angle.

**PREFACE****The Return**

The thesis, the reading contract, and the journey through the full eight-part argument.

**PART 1****Echoes**

Dotcom memory, paradigm shifts, and why real technologies produce the most dangerous bubbles.

**PART 2****Prototypes**

Early companies and interfaces are prototypes of futures they may not own. A prototype can be directionally right and economically doomed.

**PART 3****Moats**

Intelligence alone is not the moat. Durable moats may be context, distribution, workflow, identity, permissions, governance, execution rights, data and systems of record.

**PART 4****SaaS Was a Compromise**

SaaS won because custom software was too expensive. AI changes that cost curve, and the compromise gets repriced.

**PART 5****Firmware**

Firm-ware is the operating logic embedded into the company itself. AI makes that logic programmable.

**PART 6****Breaking Point**

The market does not stop believing in AI. It starts doubting who owns it.

**PART 7****After the Crash**

The crash is a repricing event. The mature AI economy begins after the first ownership story fails.

**PART 8****The New Power Map**

A layered view of the durable winners, pressured groups, and control points in the mature AI economy.

## How to read the series

Each chapter works on its own, but this is designed as a sequential essay.

Read straight through and the structure should feel intentional:

- [Part 1](#) returns to dotcom memory and the emotional pattern of paradigm-shift bubbles.
- [Part 2](#) states the core claim directly, AI is useful, and that is exactly why the bubble is dangerous.

- [Part 3](#) asks whether intelligence delivered as an API is a durable moat or only an early distribution phase.
- [Part 4](#) argues that SaaS benefited from the old cost of building software, and AI changes that cost curve.
- [Part 5](#) explores the AI-operated company as a possible new organizational form.
- [Part 6](#) maps the pressures that could cause belief to break.
- [Part 7](#) asks what becomes more real after the crash, not what disappears.
- [Part 8](#) ends with the likely power map, who is strengthened, who is pressured, and where durable control may live.

## Start with Part 1

If you want the series in the intended order, begin with [Part 1: The Business Cards in the Drawer](#).

That opening chapter is where the memory, mood, and pattern first lock together. Everything after that builds from the same intuition:

*the technology can be real, the excitement can be justified, and the market can still be pricing the wrong future.*

If that premise interests you, the journey map above is your way in.

PART 1

# Echoes

---

## Part 1: Echoes

### The business cards in the drawer



I still have two packs of business cards from people I met at investor and founder events in San Francisco around 2000.

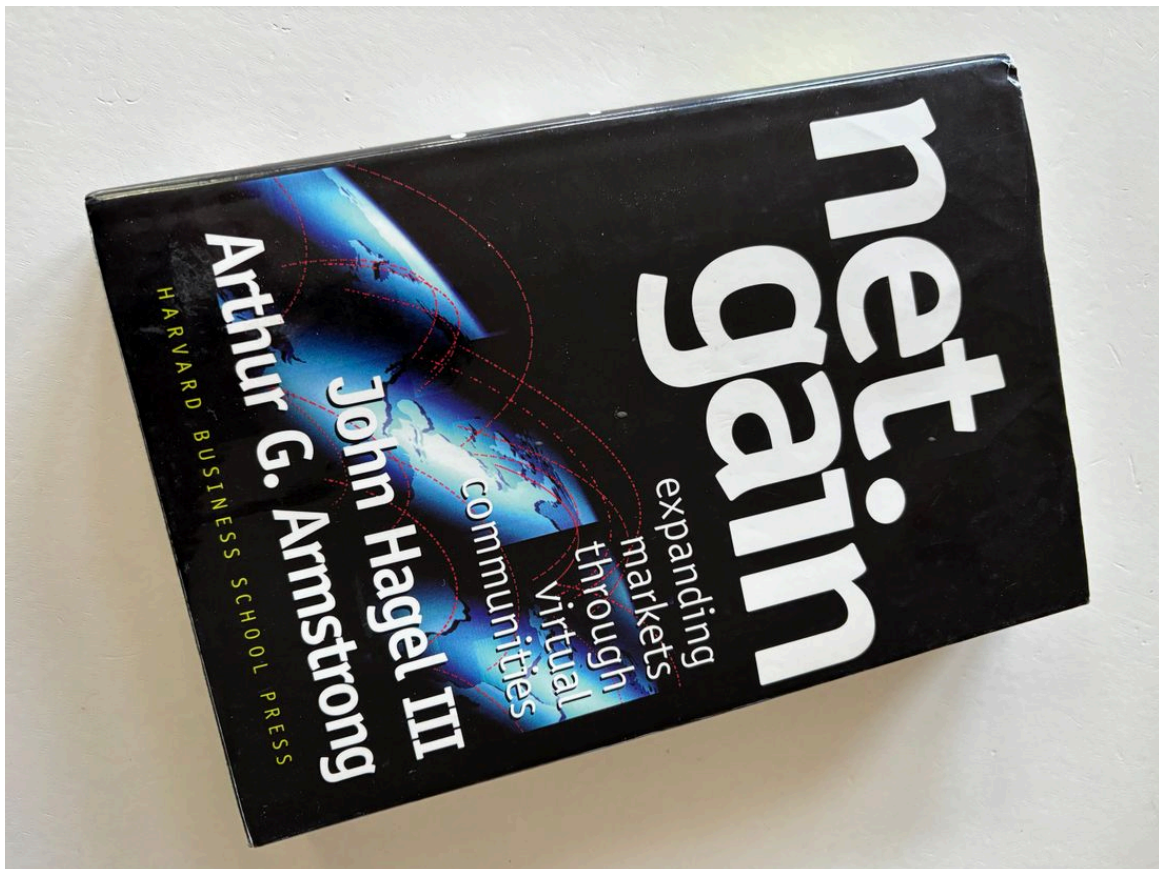
I kept them for posterity.

At the time, I did not know quite what I was keeping. I was working with a company inside the dotcom bubble. I spent time around people seeking investment, pitching ideas, trying to explain why the internet would change everything, and why their particular company would be one of the companies that survived and captured the upside.

There was a strange atmosphere around it all. It was not just greed, although there was plenty of greed. It was not just naivety, although there was plenty of that too. It was a feeling that something genuinely enormous was happening and that everyone in the room was close enough to touch it.

The future felt visible. Not clear. Not mature. Not understood. But visible.

## Net Gain and the future hiding in plain sight



One of the books that seemed to follow people around the tech and startup scene in San Francisco at the time was [Net Gain: Expanding Markets Through Virtual Communities by John Hagel and Arthur Armstrong](#). It was published in 1997, by Harvard Business School Press, and in the circles I was moving through it had something close to cult status. Not cult in the mass-market sense. Cult in the way certain books move through a scene because they give people language for something they can already feel but cannot yet fully explain. This book was giving the startup and tech communities a blueprint for the future. It described how e-commerce and communities were going to become central to the economy. This was subliminal and nobody had really drawn out such a clear map before.

What made Net Gain strange in hindsight is that it was, in effect, describing social networks before we had the phrase “social network” in the way we use it now. It was not describing Facebook, LinkedIn, Reddit, Twitter or the platform economy directly. Those things came later. But it understood something fundamental: that the internet was not only a publishing medium, a shopping channel or a cheaper distribution mechanism. It was a place where people would gather, identify with communities, exchange information, build trust, form preferences and create new markets around shared interest. Wired captured the mood at the time, noting that the book had been handed around at “digerati summits” and had travelled through the scene by word of mouth.

That is what I mean when I say the future felt visible but not clear. Net Gain was not right about every detail, and the businesses that tried to act on that idea in the late 1990s often had the wrong timing, wrong cost structure or wrong monetization model. But the basis of the argument was largely correct. Communities did become markets. Identity, trust, recommendation, reputation, belonging and network effects did become core economic infrastructure. The book saw the outline of the mature internet while the technology was still in its adolescence.

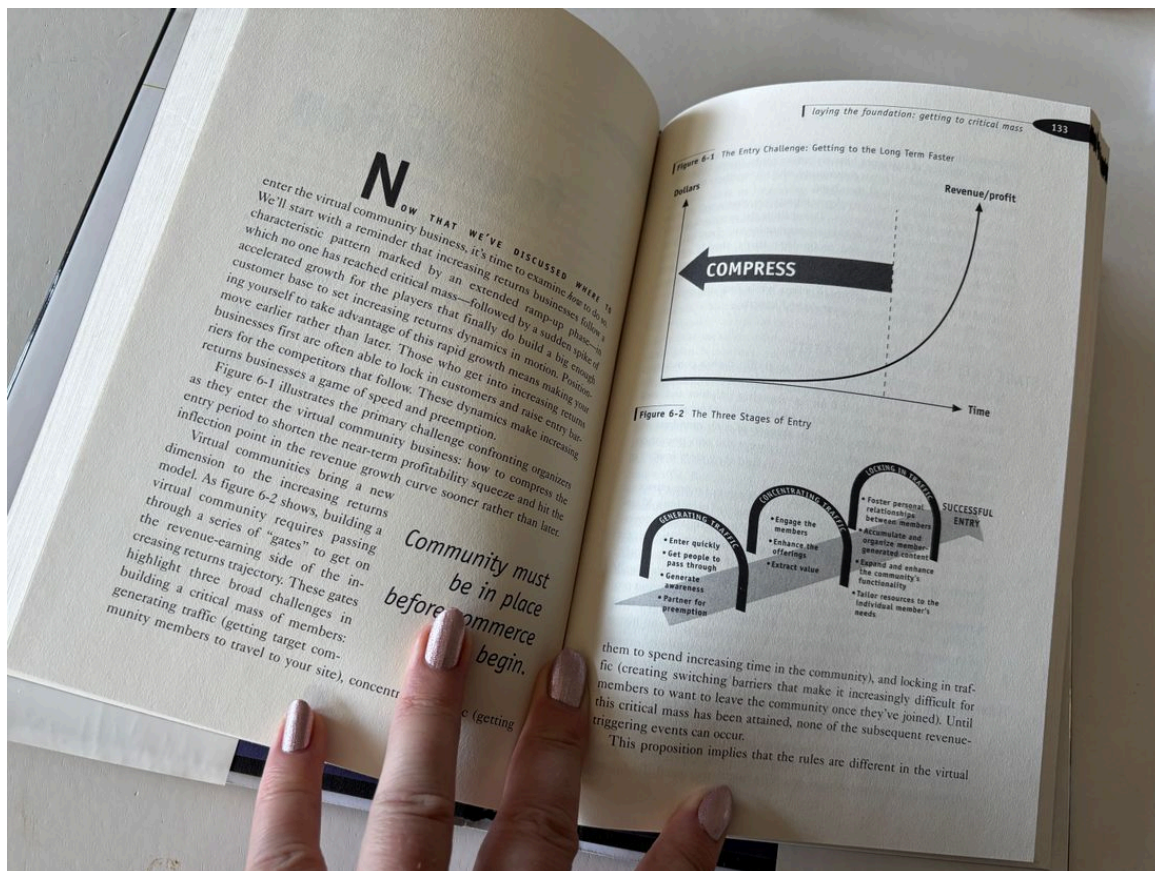
This is exactly the kind of thing that makes a paradigm-shift bubble so hard to reason about from inside it. The early language sounds overexcited because it is reaching for something that does not properly exist yet. Some of the companies built around that language will be ridiculous. Some will disappear. Some will burn absurd amounts of capital. But underneath the bad timing and the failed companies, there can still be a real prediction trying to get out.

You could feel the shape of something. You could feel that the internet was going to change commerce, communication, media, work, identity, finance, software and culture. The mistake was not believing that. The mistake was believing that this first generation of companies, with their first generation of business models, cost structures, burn rates and investor decks, already knew how that future would work.

I went there as a technical consultant for an M&A, but the acquirer were instead impressed with me enough to ask me to move to Silicon Valley and work for them. Sometimes I wonder what my life would have been like had I said yes, but that was not my story and I stayed in London. Leaving the UK would take another 4 years.

Well, I'm going to be bold and maybe a little arrogant and say, I want this to be my Net Gain moment for AI. I can feel the shape of something real. I can feel that AI is going to change commerce, communication, media, work, identity, finance, software and culture. The mistake would be to believe that this first generation of companies, with their first generation of business models, cost structures, burn rates and investor decks, already knows how that

future will work. There aren't many benefits to getting old, but one of them is recognizing patterns over a long period of time. I have seen this movie before. I know the emotional texture. I know the pattern of belief and disillusionment. I know the way the future feels obvious in hindsight, but is still so hard to get right in the moment.



## Most of the companies represented in those business cards no longer exist.

That is not a small detail for me. It is the physical memory of a market story that collapsed. Names, logos, job titles, phone numbers, clever domains, all printed with confidence on expensive card stock. Nearly all of them are gone.

But the internet did not disappear.

That is the part people forget when they talk about bubbles. They talk about the stupidity of the companies that failed, but they forget how much of the failed belief later came true in a different form.

The dotcom bubble did not prove that the internet was fake. It proved that the market had correctly identified a paradigm shift and then mispriced almost everything around it.

By paradigm shift, I mean more than a new product category or a fashionable new tool. I mean a change in the underlying logic of a system: what becomes cheap, what becomes scarce, where the power sits, and which kinds of companies eventually own the mature market.

*A paradigm shift is a fundamental, transformative change in the basic concepts, practices, and underlying assumptions of a field, system, or society. It occurs when the usual, accepted way of thinking or doing something is replaced by a new, different model, often driven by new information or technological advancements.*

That distinction matters, because I think we are now living through the same kind of moment with AI.

Not the same technology. Not the same market. Not the same infrastructure. Not the same companies.

But the pattern feels familiar:

- the same emotional texture;
- the same feeling that something real has arrived before anyone understands its final form;
- the same rush to price the mature version of a technology while it is still early, awkward and incomplete;
- the same confusion between believing in the paradigm and believing in every company claiming to own it;
- the same danger that the crash, when it comes, will be misread as proof that the whole thing was fake.

It will not be fake. That is the point. The AI bubble will burst because AI is real.

## **The internet was not wrong. The first market story was wrong.**

It is easy to laugh at the dotcom bubble now.

Pets.com became the punchline. A sock puppet, a Super Bowl advert, and a business model that seemed ridiculous in hindsight. People still use it as shorthand for everything absurd about that period.

But step back from the company and look at the idea.

Was selling pet supplies online absurd? No, Chewy exists. Amazon sells pet supplies. Supermarkets deliver pet food. Subscription commerce exists. Heavy, bulky, recurring household purchases are now a perfectly normal part of online retail.

The absurdity was not the concept. The absurdity was the timing, the economics, the logistics, the customer acquisition cost, the infrastructure, the maturity of consumer behaviour, and the belief that a first-wave company could spend its way through all of those missing pieces quickly enough to justify the valuation.

Pets.com was not wrong in the simple way people like to think. It was directionally right and structurally early.

That is often more dangerous than being wrong.

A bad idea can die cleanly. A good idea arriving too early can absorb almost unlimited capital, because every piece of evidence against it can be reinterpreted as temporary friction.

The same was true of online clothes shopping.

In the late 1990s and early 2000s, the objection seemed obvious. You cannot try clothes on through a screen. Sizing is inconsistent. Colours are unreliable. Returns are annoying. People want to feel fabric. Fashion is personal. Shops are social.

All of that was true. It just was not **permanently** true.

Over time, the missing infrastructure formed around the idea. Broadband improved. Product photography improved. Mobile commerce arrived. Payment systems became trusted. Reviews became normal. Free and easy returns changed behaviour. Warehouses became more sophisticated. Delivery became faster. Consumers became comfortable buying things they had never touched.

Now Zalando exists. ASOS exists. Fashion ecommerce is normal.

The sceptics were not stupid. They were looking at an early, awkward version of a new market and asking whether it could already do the work of a mature one.

The market made the opposite mistake. It looked at the mature destination and priced it into companies that were still learning to walk.

This is the pattern I want to describe in this series.

Not every bubble is the same. Some bubbles are mostly fraud. Some are mostly leverage. Some are mostly speculation. Some are collective madness wrapped around a thin asset.

But some bubbles happen because the world has genuinely changed and the market cannot yet work out where the value will settle.

Those are the dangerous ones. Those are the ones that can be both **stupid and right**.

## 2008 felt different

The financial crisis of 2008 did not feel like dotcom to me. It had a different texture.

The dotcom bubble was not clean or innocent, but its emotional centre was technological possibility. The internet was obviously useful. The question was not whether it mattered. The question was how, when, and who would make money from it.

2008 was different. It was a balance sheet crisis, a leverage crisis, a fraud crisis, a ratings crisis, a broken incentives crisis. It was built on opacity and mispriced risk. It was not society discovering a new general-purpose technology. It was finance eating itself and then handing the bill to everyone else.

That distinction matters because people use the word bubble too loosely.

If everything overvalued is just a bubble, we lose the ability to distinguish between a fake story and an early story.

Dotcom was an early story. AI is an early story.

Crypto is more complicated. I do not think it should be dismissed entirely. It may still become a genuine paradigm shift if it eventually displaces fiat currency, settlement systems or trusted intermediaries at scale. But so far, much of its mainstream economic life has been speculation, regulatory arbitrage, niche settlement, corruption, fraud, and early utility in grey or black markets. That does not mean nothing important can come from it. It does mean that its final form is still unproven.

AI feels different. The utility is already obvious.

Developers use it. Writers use it. Students use it. Analysts use it. Designers use it. Customer support teams use it. Researchers use it. People who do not care about technology use it.

That does not mean the valuations are sane. It means the bubble is more dangerous.

The easiest bubbles to dismiss are the ones built on fantasy. The hardest bubbles to survive are the ones built on something true.

## Society can feel the future before it understands the business model

Every rare technological shift has a strange early period, but the pattern is usually recognizable in hindsight.

It often moves through something like this:

1. **Pioneer phase:** the technology arrives looking like a toy, a curiosity, or a niche obsession.
2. **Discovery phase:** people start to feel the implications before they can explain the business model.
3. **Speculation phase:** capital, talent and language rush in ahead of settled economics.
4. **Overbuild phase:** the market starts pricing a mature economy before the infrastructure, habits and profit pools exist.
5. **Belief break:** the first story stops explaining reality well enough to support the valuations.
6. **Consolidation and maturity:** the technology keeps spreading, but ownership shifts and the real winners often emerge later.

It happened with railways. It happened with electricity. It happened with computing. It happened with the internet. It happened with mobile. It happened with cloud.

The first version of the story is almost always immature.

The first railway mania was not the final railway economy. The first electrical companies were not the final electrical grid. The first personal computer companies were not the final computing industry. The first web portals were not the final internet. The first mobile apps were not the final mobile economy. The first cloud startups were not the final cloud platform structure.

The first story is usually a cartoon version of the final story.

But cartoons matter.

They compress the future into something people can point at. They create the first wave of capital, talent, language, infrastructure and public imagination.

The problem is that markets do not merely imagine. They price.

And once the market starts pricing phase six while the world is still somewhere around phases two, three or four, the bubble begins.

AI is at exactly this stage.

People can feel the future. They can feel that software is changing. They can feel that work is changing. They can feel that knowledge work, creativity, analysis, coding, support, operations, education and business process design are all becoming fluid in a way that was not true before.

They are right to feel that.

But the market is already trying to decide who owns the future.

- Is it the model labs?
- Is it the hyperscalers?
- Is it the chip companies?
- Is it the SaaS incumbents adding AI features?
- Is it the device and operating system companies?
- Is it the companies with proprietary data?
- Is it whoever owns identity, permissions, workflow, governance and orchestration?
- Is it a new class of AI-operated companies that do not yet exist at scale?

The answer matters because capital markets do not reward vague transformation. They reward ownership, pricing power, margins and durable control.

That is where I think the current story begins to fracture.

## **The mistake is confusing the gateway with the destination**

In the early internet, gateways mattered enormously.

For many people, AOL was the internet. It was access, identity, email, chat, content, community, billing and a sense of safety. It translated the chaos of the open web into something ordinary people could use.

That was historically important.

But being the gateway to a paradigm in its childhood does not guarantee ownership of the mature paradigm.

The mature internet was not AOL.

It became search, broadband, cloud, social networks, ecommerce, mobile operating systems, app stores, streaming, online advertising, creator platforms and walled gardens.

The value migrated.

Some of the early companies survived. Many did not. Some infrastructure was overbuilt and later reused. Some companies became footnotes. Some assets were absorbed. Some ideas reappeared years later with better timing.

The public story changed.

That is the lens through which I see ChatGPT.

ChatGPT made AI legible to normal people. That is an enormous achievement. For many people, ChatGPT is AI in the same way AOL was once the internet.

But that does not mean a chatbot interface, or a centralized model API, is the final economic structure of AI.

It may be the gateway.

The mature form may be something much more distributed, embedded, local, specialized and invisible.

This is where the current AI market story becomes fragile.

A great gateway can create a category. It does not automatically own the category forever.

## **A paradigm shift is not a company**

One of the most important lessons of dotcom is that a paradigm shift and the companies funded around that paradigm are not the same thing.

- The internet was real. Many internet companies were not durable.
- The internet transformed commerce. Many ecommerce companies failed.
- The internet transformed media. Many media startups failed.
- The internet transformed software. Many early software-as-a-service companies failed or were absorbed.
- The internet transformed advertising. Many adtech companies were replaced, consolidated or turned into infrastructure for larger platforms.
- The internet transformed almost everything, but not in the way most pitch decks described at the time.

This is the warning for AI.

- AI can be a real paradigm shift and OpenAI/Anthropic can still be mispriced.
- AI can transform business and Anthropic-style model API economics can still be overestimated.
- AI can make companies more productive and SaaS incumbents can still lose pricing power.
- AI can create enormous demand for compute and some data centre investments can still turn out to be badly timed.
- AI can change software forever and thin AI wrappers can still vanish.
- AI can be the future and investors can still lose money funding the wrong layer.

That last sentence is the whole point.

A paradigm shift does not tell you where the profit pool settles.

It tells you that the map is being redrawn.

## Why the current AI bubble feels familiar

The current AI moment has the same strange combination of truth and overreach that I remember from dotcom.

The truth is obvious.

AI is useful. It is not just a demo. It is not just a toy. It is not just autocomplete. It is not just a better search box.

Used well, it changes the cost and speed of producing text, code, analysis, images, summaries, plans, prototypes and decisions. It compresses the distance between intent and output.

That matters, but then comes the overreach.

- Because AI can write code, people assume software development as an economic activity is about to collapse into a few prompts.
- Because AI can answer questions, people assume enterprise knowledge management is solved.
- Because AI can produce plausible plans, people assume autonomous agents can run complex organizations without supervision.
- Because companies are spending heavily on AI infrastructure, people assume demand will compound smoothly enough to justify every data centre, every GPU order and every financing structure.
- Because a model is best today, people assume it has a moat tomorrow.
- Because a SaaS company adds AI features, people assume it has defended itself.
- Because a startup calls itself AI-native, people assume it belongs to the future rather than to the first wave of overfunded experiments.

That is how bubbles form around real technologies.

Reality provides enough evidence to make the fantasy plausible.

## The market is pricing the mature AI economy before we know its shape

The mature AI economy will not look like the first AI economy.

That is the safest prediction I can make.

The first AI economy is visible now:

- model labs racing for frontier capability;
- hyperscalers building enormous data centre capacity;
- chip companies selling into the buildout;
- SaaS companies bolting AI features onto existing products;
- enterprises running pilots;
- startups wrapping APIs;
- consultants selling transformation;
- investors looking for the next platform shift.

Some of this will survive. Some of it will become foundational. Some of it will be absorbed. Some of it will become embarrassing.

The mature AI economy may look very different.

Models will likely become smaller, cheaper, more specialized and more local. The best frontier models may still matter enormously, but not every task will need to call a giant remote model. Much intelligence will move closer to the device, the browser, the operating system, the enterprise tenant, the application and the workflow.

Software itself will become more fluid. Today, companies buy SaaS because custom software is expensive, slow and risky. If AI makes custom software cheaper, faster and more maintainable, then large parts of SaaS begin to look like a historical workaround rather than a permanent final form.

Organizations will change too. The most radical AI businesses may not be companies that sell AI. They may be companies that are mostly run by AI. Tiny human teams, large automated surfaces, generated internal software, AI-driven support, AI-driven operations, AI-driven reporting, AI-driven research, AI-driven sales workflows and continuous adaptation.

This is not a distant philosophical idea. It is the direction of travel once software creation, workflow execution and business coordination all become cheaper at the same time.

That is why the capital markets story matters.

The market is not simply betting that AI is useful. It is betting, implicitly, on where the economic control points will be.

I think many of those bets are wrong.

## **The crash is a change in belief**

Bubbles do not burst the moment something becomes overvalued. They can stay overvalued for a long time. They burst when belief changes.

That belief change is rarely neat. It does not require everyone to wake up on the same morning and decide the future is cancelled. More often, the story loses its ability to explain contradictory evidence.

In dotcom, the internet did not stop being important. What changed was belief in the particular companies, burn rates and business models that had been funded.

At some point, investors no longer believed that every company with a domain name was a future monopoly.

- They no longer believed that traffic alone was enough.
- They no longer believed that losses were automatically evidence of ambition.
- They no longer believed that first-mover advantage could overcome bad unit economics.
- They no longer believed that capital could buy time forever.
- The future was still coming, but the story had broken.

That is what I expect in AI.

The bubble will not burst because people stop using AI. It will burst when enough people realize that the current story about who owns AI is wrong.

The convergence point may be a combination of pressures:

- model capability improves, but not fast enough to justify every valuation;
- smaller, local and open models reduce the scarcity premium of centralized APIs;
- enterprises discover that AI value is real, but integration is slower and harder than the demos implied;
- token costs become visible once subsidized and flat-rate usage gives way to harder economics;
- hyperscaler capex becomes harder to justify at the same growth rate;
- SaaS companies see pricing pressure, seat pressure and growth pressure;

- thin AI wrappers fail;
- investors ask who really owns the future.

The technical story does not have to fail. The market story only has to become unbelievable.

## **After the crash, the real economy begins**

The strangest thing about the dotcom crash is that, in hindsight, it looks less like the death of the internet and more like the beginning of its mature life.

The crash cleared out companies, reset valuations, consolidated assets, disciplined business models and created the conditions for the next phase.

The internet after the crash became more useful, more profitable and more central to everyday life.

It also became more centralized, more extractive and more controlled.

The early web was weird, open, independent and chaotic. The mature web became search monopolies, social platforms, cloud platforms, app stores, walled gardens, algorithmic feeds, advertising markets, surveillance capitalism and the consumer as the product.

The mature internet was not the utopia many people imagined. It was better in some ways, worse in others, and much more powerful than the bubble version.

AI may follow the same pattern. The crash will not be the end of AI. It may be the moment AI becomes serious.

After the crash, weak model labs may disappear or be absorbed. Infrastructure will be reused. Inference will become cheaper. AI will become embedded into operating systems, devices, browsers, enterprise platforms and workflows. SaaS will consolidate. Custom workflow software will become more viable. AI-operated companies will become real competitors. The power map will shift toward chips, devices, private data, identity, governance, orchestration and systems of record.

The social consequences will be enormous. Labour markets, education, creative work, software development, management, services and the structure of companies themselves will all be affected. That deserves its own essay.

This series is narrower. It is about business, capital markets and the mistake of pricing the wrong ownership layer.

## The first story fails

When I look at those old business cards, I do not see proof that everyone was stupid.

I see proof that people can be close to a real future and still be wrong about almost every investable expression of it.

That is the lesson.

The internet was real. The first story failed. AI is real and I believe the first story will fail again.

- Not because AI is useless.
- Not because the demos are fake.
- Not because nothing will change.

But because everything will change, and the market has not yet understood where that change finally settles.

The AI bubble will burst because AI is real.

That is where this series begins. I hope you'll enjoy this journey with me into my vision of the future.

## PART 2

# Prototypes

---

## Part 2: Prototypes

[Part 1](#) argued that markets routinely misprice real paradigm shifts by pricing the mature economy too early. This chapter asks the harder follow-up question: if AI utility is real, where does the value actually settle, and who captures it before the story breaks again in [Part 3](#)?

The early companies and interfaces that make a new paradigm legible are prototypes of futures they may not own.

*A prototype can be directionally right and economically doomed.*

That is the central claim of this chapter. Pets.com understood something real about the internet. AOL made the web accessible to millions. The first AI chat applications are teaching ordinary people to think in prompts. Each of those is a prototype: evidence that the direction is correct, not proof that the current owners are the final ones.

The easiest way to misunderstand the AI bubble is to assume there are only two possible positions.

Either AI is revolutionary, therefore the valuations make sense, or AI is overvalued, therefore the technology is fake.

I think both positions are wrong. The technology is real. The valuations can still be wrong.

That was the first lesson of the dotcom bubble. A technology can be world-changing and still destroy enormous amounts of capital in its first speculative wave. The internet did not fail in 2000. The market story failed. Investors had correctly sensed a paradigm shift, then funded almost every early expression of that shift as if it had already reached its mature form.

That is why the current argument around AI often feels so shallow. People keep asking whether AI is real, whether it is useful, whether it is better than search, whether it can code, whether it can reason, whether it can replace workers, whether companies are using it.

Those questions matter, but they are not enough.

The more important question is this:

*If AI is real, who actually captures the value?*

That is where the bubble sits.

Not in the claim that AI matters.

In the claim that the companies currently priced as if they own the future will, in fact, own it.

## **Real utility creates more convincing bubbles**

The most dangerous bubbles are not built on nothing.

They are built on enough truth to make the overreach feel rational.

AI has that truth.

[Stanford's AI Index](#) is one of the clearest high-level reminders that AI is no longer a speculative abstraction. And Microsoft now markets [Microsoft 365 Copilot](#) as something woven directly into the everyday work surface of documents, meetings, files and chats. Those facts do not prove the current valuations are correct. They do make the bubble more convincing.

A developer using a coding assistant can feel it. A student using it to understand a topic can feel it. A writer using it to structure a draft can feel it. A support team using it to summarize tickets can feel it. A manager using it to synthesize documents can feel it. A designer using it to explore visual directions can feel it. A small business owner using it to write copy, build a spreadsheet or understand a contract can feel it.

This matters.

AI is not a pitch deck technology. It is already in people's hands. The usage is not theoretical. The productivity gain is uneven, but it is not imaginary.

That is exactly why the bubble can run so far.

When people can feel a technology working, they become much more willing to believe that the current winners are inevitable. They confuse personal proof of utility with market proof of durable economics.

I can use AI every day and still believe OpenAI is overvalued.

I can believe AI will transform software and still believe many SaaS companies will be repriced downward.

I can believe AI will become part of almost every business and still believe that much of the current infrastructure spending will be mistimed, overbuilt or monetized by someone other than the people funding it.

This is the central paradox of the series:

*AI is real enough to destroy the current AI trade.*

## **The market is pricing too many assumptions at once**

The current AI market story rests on several assumptions.

First, that frontier models remain scarce and expensive enough for a few model labs to capture large rents.

Second, that hyperscaler capital expenditure converts into durable, highly utilized, high-margin AI revenue.

Third, that SaaS incumbents can add AI features and preserve their existing application and workflow rents.

Fourth, that enterprise adoption moves from experimentation to deep operational transformation quickly enough to justify the spend.

Fifth, that AI agents become reliable enough to automate meaningful amounts of work, while still depending on centralized cloud model infrastructure.

Sixth, that customers tolerate the pricing required to support the infrastructure.

Seventh, that the value of AI does not rapidly commoditize as models get smaller, cheaper, open, local and embedded.

That is a lot of belief stacked together.

A bubble does not need every assumption to be false. It only needs enough of them to soften at the same time.

The market may be broadly correct that AI is a paradigm shift and still be wrong on timing, winners, margins, architecture, pricing power and capital intensity.

That is how real-technology bubbles work.

## **This is not another 2008**

It is useful to separate this from 2008.

The financial crisis was not society over-extrapolating a new general-purpose technology. It was a crisis of leverage, opacity, securitization, ratings, incentives and fraud. The system pretended that risk had been transformed when much of it had simply been hidden, sliced, resold and misunderstood.

The dotcom bubble had bad behaviour too. Every bubble does. But its emotional centre was different. It was not primarily a hidden balance-sheet rot. It was a visible overbelief in a new technology.

AI feels closer to dotcom than to 2008.

Not because the assets are identical. They are not.

Not because the same companies are involved. They are not.

Not because the crash will look the same. It will not.

It feels similar because the psychological structure is similar.

People can see the future, but they cannot yet see the final business model.

So the market invents one.

In the late 1990s, the invented story was that every business needed to become a dotcom and that first movers with traffic, brand and capital would dominate huge new markets.

Today, the invented story is that intelligence will flow through a small number of model and cloud platforms, that every SaaS incumbent can defend itself by adding AI, and that enterprise adoption will generate enough cash flow to justify the infrastructure buildout.

Some of that will be true.

Not enough of it has to be true for every valuation to survive.

## **The wrong lesson from Pets.com**

People use Pets.com as if it proves that everyone in the dotcom bubble was delusional.

I think it proves something more interesting.

The idea was early, not permanently absurd.

A pet supply business on the internet is normal now. Subscription pet food, online pharmacy, bulk delivery, automatic repeat orders, logistics optimization, customer reviews, card payments, warehouse networks, mobile commerce, all of that now exists around the category.

The early company failed because the world around the idea had not matured.

The same lesson applies to AI.

Some ideas that look absurd now will become normal later. Some current AI companies that look central now will vanish. Some use cases that feel clumsy today will become invisible infrastructure tomorrow. Some infrastructure built too early may become valuable later under different ownership.

The market will not be wrong to believe in AI.

It will be wrong to believe that the first visible structure is the final structure.

## **Utility does not equal capture**

This is the most important distinction.

Utility is not capture.

A technology can create enormous value for society while the companies that funded the first wave capture less of it than expected.

The internet created enormous consumer surplus. It made communication, search, publishing, commerce and software distribution radically cheaper. But many of the first companies that created or popularized those behaviours did not survive.

The value migrated.

It migrated to search. It migrated to cloud. It migrated to social networks. It migrated to app stores. It migrated to mobile operating systems. It migrated to advertising platforms. It migrated to payment systems. It migrated to logistics networks.

AI value will migrate too.

It may migrate away from pure model APIs toward chips, devices, local inference, operating systems, browsers, enterprise data, identity, workflow orchestration, systems of record and AI-operated businesses.

It may migrate away from generic SaaS applications toward custom generated software and internal control planes.

It may migrate away from human-heavy services companies toward tiny AI-operated competitors.

The AI economy can grow while many AI-era valuations collapse.

That is not a contradiction.

It is the most likely path for a real paradigm shift.



*A paradigm shift can feel tangible on the desk long before the ownership map is clear.*

---

**EDITORIAL FIGURE**

## Utility is real. Capture migrates.

The useful technology and the durable ownership layer do not appear at the same moment. That timing gap is where the bubble hides.

---

**THESIS LINE**

*AI is real enough to destroy the current AI trade.*

---

### What people can feel

Better drafting, coding, summarising, research, support and decision support.

---

### What markets overclaim

That the first visible leaders, pricing models and capital structures must own the mature economy.

---

**VALUE MIGRATION**

---

**01 Utility becomes obvious**

People can feel AI working in everyday tasks long before the economics settle.

---

**02 The first story overclaims**

Markets mistake visible usefulness for proof that today's winners own the mature profit pool.

---

**03 Value starts migrating**

Capture shifts toward workflow, governance, devices, records, orchestration and execution.

---

**04 Durable ownership forms later**

The real profit pool settles where intelligence meets authority, trust and operational control.

---

**WHERE DURABLE CAPTURE GOES**

Devices and operating systems

Systems of record

Identity and governance

Workflow orchestration

Execution control planes

AI-operated businesses

## **Why enterprise adoption will be slower and deeper than the demos suggest**

AI demos are seductive because they remove the mess.

A clean prompt. A clean answer. A generated app. A summarized document. A working prototype. A sales email. A graph. A plan. A piece of code.

But businesses are not demos.

Real businesses have permissions, policies, compliance requirements, legacy systems, bad data, partial data, missing data, political constraints, audit requirements, human exceptions, edge cases, incentives, vendor contracts, customer commitments and regulators.

This does not make AI less important.

It makes the transformation slower, deeper and more structural than the market wants to price.

The real value of AI in an enterprise is not a chatbot sitting next to a process. It is AI embedded inside the process, with access to the right data, governed by the right permissions, orchestrated through durable workflows, evaluated against real outcomes and constrained by human approval where needed.

That is hard.

It is not impossible. It is precisely where the long-term value sits. But it means a lot of early enterprise AI spending will disappoint before the real operating model emerges.

Again, this is dotcom logic.

A website was not a business model.

A chatbot is not a business model.

## **The first AI story will fail**

I think the first AI story is already visible.

It says that the model labs own intelligence. The hyperscalers own the compute. The SaaS companies own the workflows. Enterprises will pay for AI features inside the software they already use. Agents will increase usage. Infrastructure spending will keep compounding. The

companies closest to today's AI excitement will become tomorrow's dominant economic layer.

That story is too neat.

The real story is likely messier.

Models will get smaller. Inference will get cheaper. More capability will move locally. Open models will keep improving. Devices and operating systems will absorb more intelligence. Enterprises will build private and hybrid AI stacks. SaaS will be pressured by custom software. AI-operated companies will compete with human-heavy incumbents. Some infrastructure will be overbuilt. Some capital will be burnt. Some of today's winners will remain winners, but for different reasons than the market currently believes.

The bubble will burst when the first story can no longer carry the valuation.

That does not require AI to fail.

It requires AI to mature.

That is why I do not think the AI bubble will burst because AI is fake.

I think it will burst because AI is real, and because reality will be more disruptive to today's market story than scepticism ever could be.

In [Part 3](#), I take that question directly into the current ownership bet: the idea that intelligence as an API is itself the durable moat.

## PART 3

# Moats

---

## Part 3: Moats

[Part 2](#) made the distinction between utility and capture. This chapter applies that distinction to the strongest current ownership claim in AI: that a few centralized model providers will remain the permanent home of intelligence. I think that story is much weaker than it looks, and it sets up the SaaS consequences explored in [Part 4](#).

Intelligence alone is not the moat.

*Intelligence will flow through a small number of centralized model providers because they have the best models and the infrastructure to run them.*

That sounds reasonable today. It may even be true today, but I do not think it is the final form.

OpenAI, Anthropic and the other frontier labs are betting on a world where intelligence remains scarce, expensive, centralized and hard to reproduce. They have the models. They have the talent. They have the infrastructure relationships. They have the APIs. They have the mindshare. They have the brand.

That is a powerful position, but it is not automatically a durable moat.

A model lead is not a moat. It is a lead, and leads decay.

## Scarcity is the current business model

The economics of the frontier labs are built around scarcity.

The best models are expensive to train. They are expensive to serve. They require huge amounts of compute, power, memory, networking and operational expertise. They sit behind APIs because very few companies can afford to train or run them at scale.

That creates the feeling of a moat.

If everyone needs intelligence and only a few companies can provide it, then those companies should capture the value.

That is the simple version of the story. It is also the version I am most suspicious of.

### **Technology attacks scarcity.**

That is what it does.

Every layer of the AI stack is currently under pressure to become cheaper, smaller, faster and more distributed. Model architecture improves. Inference improves. Quantization improves. Distillation improves. Hardware improves. Memory improves. Edge devices improve. Developer tooling improves. Open models improve. Specialized models improve.

- The question is not whether frontier models will remain impressive. They will.
- The question is how many tasks actually require the frontier.

That number may be much smaller than the market currently assumes.

## **Yesterday's frontier becomes tomorrow's local model**

Computing history repeatedly moves capability from centralized infrastructure to local devices.

Mainframes mattered. Then personal computers mattered. Then servers and cloud mattered. Then mobile and edge mattered. The pattern is not a clean replacement. It is a migration of the default.

The same thing is likely to happen with AI.

Today's expensive cloud capability becomes tomorrow's cheap local capability. Not all of it. Not the moving frontier. But enough of it to change the economics.

A local model does not need to be the smartest model in the world to be valuable.

It needs to be good enough for the task, cheap enough to run continuously, private enough to trust, fast enough to feel ambient and integrated enough to disappear into the workflow.

That is a different design target from the frontier leaderboard.

Most everyday AI tasks are not grand acts of genius.

They are repetitive, contextual and close to the user:

- summarize this thread;
- rewrite this paragraph;
- classify this email;
- extract these fields;
- explain this error;
- draft this reply;
- organize these notes;
- search my files;
- help with this form;
- suggest the next step;
- generate a small script;
- automate this local workflow;
- check this document against a known policy.

Those are not all frontier tasks.

They are context tasks.

The advantage goes to whoever owns the context, the device, the operating system, the workflow, the permissions and the user relationship.

That is often not the model lab.

## **The future is hybrid, not API-only**

I am not arguing that cloud AI disappears.

That would be too simplistic.

The frontier will matter. There will be tasks where the best available model is worth paying for. Complex reasoning, scientific work, high-end coding, multimodal generation, long-context synthesis, difficult planning, regulated review, specialist analysis and heavy agentic workflows may all require cloud-scale models for a long time.

But the default will shift.

The mature AI stack is likely to be hybrid:

- local models for private, low-latency, ambient and repetitive tasks;
- device and operating system models for personal context;
- enterprise-local or tenant-local models for sensitive business data;
- specialized models for narrow domains;
- frontier cloud models for escalation;
- orchestration layers deciding what runs where.

Apple's [Apple Intelligence](#) is already explicitly framed around on-device processing with Private Cloud Compute, and Microsoft is using [Copilot+ PCs](#) to push local NPU-based AI further down into the device. That does not prove the hybrid future is settled. It does show that the market is already moving beyond the fantasy that every meaningful act of intelligence stays behind one remote endpoint.

That is very different from a world where every meaningful unit of intelligence flows through a handful of remote APIs.

The API remains useful.

It stops being the entire architecture.

**FIGURE****API-only versus hybrid AI**

The current centralized market story versus the hybrid stack this chapter argues toward. The point is not that the API disappears — it is that it stops being the whole architecture.

**API-only story**

A few remote model providers remain the default path for nearly all meaningful intelligence.

**TAKEAWAY**

This is the current market story.

**DEFAULT SURFACE**

Users go to one destination product or API endpoint to reach intelligence.

**WHERE CONTEXT LIVES**

Important context must be shipped outward to the model provider.

**ECONOMIC LOGIC**

Scarcity and API access look like the main source of durable rents.

**WHAT BECOMES DURABLE**

The moat is assumed to sit mainly in frontier model quality and infrastructure access.

**Hybrid / local future**

The mature stack routes work across local, embedded, tenant-local, and frontier systems based on context and cost.

**TAKEAWAY**

This is the mature-economy architecture the essay is pointing toward.

**DEFAULT SURFACE**

Intelligence sits inside the editor, device, operating system, enterprise control plane, or workflow itself.

**WHERE CONTEXT LIVES**

More context stays close to the user, the tenant, the system of record, or the device.

**ECONOMIC LOGIC**

Model access still matters, but value shifts toward orchestration, permissions, distribution, and execution.

**WHAT BECOMES DURABLE**

The moat sits where intelligence meets workflow, authority, defaults, and customer relationship.

**Distribution beats model purity**

In immature markets, people overvalue the pure technology layer.

They assume the best technology captures the market.

Sometimes it does.

Often it does not.

Distribution matters. Defaults matter. Workflow matters. Trust matters. Data matters.  
 Integration matters. Switching costs matter. Procurement matters. Regulation matters.  
 Developer ecosystems matter. Existing user behaviour matters.

<b>Claimed moat</b>	<b>Why it looks strong today</b>	<b>Why it weakens in the mature AI economy</b>	<b>What becomes more durable</b>
Frontier model quality	The best model is still visibly better	Many commercial tasks only need good-enough intelligence	Context, workflow, permissions, user trust
Centralized API access	It is the easiest way to ship intelligence quickly	More tasks become local, embedded, hybrid, or routed	Orchestration and distribution
Brand mindshare	Users currently go to the model destination on purpose	Mature AI becomes ambient and less model-visible	Default surfaces inside devices and operating systems
Training scale alone	Scarcity looks like a moat in the boom phase	Scarcity gets attacked by diffusion, distillation, and specialization	Control of execution and customer relationship

That is why device and platform companies are so important.

Apple can put models on the device, inside the operating system, close to photos, messages, mail, calendar, files, apps, identity and privacy controls.

Microsoft can put models inside Windows, Office, Teams, GitHub, Azure, identity, security tooling and enterprise administration.

Google can put models inside Android, Search, Chrome, Workspace, Gmail, YouTube, Maps and cloud infrastructure.

Samsung, Qualcomm, AMD, Intel and others can participate through devices, NPUs and local compute.

The model lab may have the better model today.

The platform company has the user, the context and the default path.

That matters because most users do not want to choose models. They want features that work.

Nobody asks which machine learning model sorted their inbox. They care whether the inbox is useful.

Nobody cares which ranking algorithm surfaced the right photo. They care that the photo appears.

Nobody cares which speech model transcribed their note. They care that the note is there.

As AI matures, model identity may become less visible to the user.

That is a dangerous future for pure model brands.

## **ChatGPT as the AOL of AI**

I do not mean this as an insult.

AOL was historically important. It made the internet accessible to millions of people. It gave people email, chat, content, billing, community and a safe front door into a confusing new world.

For many people, AOL was the internet.

Until it was not.

The mature internet did not belong to AOL. It belonged to a different set of layers: broadband, browsers, search, ecommerce, cloud, mobile operating systems, social networks, app stores, advertising platforms and streaming services.

ChatGPT may play a similar role.

It made AI legible. It gave ordinary people a way to touch the new paradigm. It turned artificial intelligence from a background technology into a direct consumer experience.

That is enormous.

But being the first widely understood gateway does not guarantee ownership of the mature economy.

The mature AI economy may not be a chat box.

It may be invisible.

It may live inside operating systems, tools, devices, workflows, enterprise control planes and generated software.

It may be everywhere and nowhere.

That would make ChatGPT historically important even if the final profit pool migrates elsewhere.

## **Model quality diffuses**

The other danger for the API moat is diffusion.

Model quality spreads faster than people expect.

The frontier may keep moving, but the trailing edge becomes stronger. Open models improve. Smaller models improve. Techniques move through papers, code, model releases, talent movement and imitation. Distillation moves behaviour from larger models into smaller ones. Synthetic data improves training. Specialized models beat general models in narrow domains.

This creates a compression effect.

The best model may remain meaningfully better at the hardest tasks.

But many commercial tasks only require good-enough intelligence.

Once good-enough intelligence is cheap, local or bundled, pricing power falls.

This does not destroy the frontier labs. It changes what they can charge for and how often they are needed.

The pure API model becomes more like premium escalation.

Useful. Valuable. Important.

Not universal.

## **Enterprise buyers will resist permanent dependency**

Large enterprises do not like unnecessary dependency.

They tolerate it when there is no alternative. They pay for Salesforce, SAP, Microsoft, ServiceNow, Oracle, Workday, Atlassian and cloud platforms because those products are deeply embedded and difficult to replace. But every dependency becomes a negotiation point.

AI increases the stakes.

Sending sensitive enterprise context to an external model provider is not a trivial architectural decision. It raises questions about privacy, security, regulation, auditability, data residency, IP leakage, vendor lock-in, cost predictability and operational control.

Some companies will use external APIs heavily.

Others will prefer tenant-local, private-cloud, self-hosted, open-model or hybrid approaches.

The more strategic AI becomes, the more enterprises will want control.

This again weakens the idea that a few centralized model APIs become the permanent default for business intelligence.

Enterprises may buy frontier capability.

But they will also build internal AI control planes.

They will route tasks. They will classify data. They will decide what runs locally, what runs in a private tenant, what runs through a managed API and what requires human approval.

In that world, the model provider is a supplier.

Not necessarily the owner of the workflow.

## **Infrastructure can be real and still overbuilt**

The AI infrastructure buildout may prove useful over time.

That does not mean every investment is well timed.

The dotcom era overbuilt parts of the internet's physical and financial infrastructure. Some of that capacity became useful later. Some of it was bought cheaply after the crash. Some investors lost everything while later companies benefited from the assets.

AI may repeat that pattern.

Data centres, chips, power contracts, cooling systems, networking, memory and specialized hardware may all be needed for the mature AI economy.

But if the market prices infrastructure as if centralized frontier inference will grow without interruption, while actual demand shifts toward local, hybrid and cheaper models, then the economics change.

The infrastructure can be useful and still disappoint the investors who funded it.

This is a recurring theme in paradigm shifts.

The asset can survive.

The capital structure cannot.

## **The model labs need to become something else**

OpenAI, Anthropic and their peers are not doomed.

That is not the argument.

The argument is that model quality alone is unlikely to be a permanent moat.

To capture durable value, model labs need to become one or more of the following:

- consumer platforms;
- enterprise platforms;
- operating layers;
- developer ecosystems;
- trusted infrastructure providers;
- device or OS partners;
- orchestration platforms;
- data and workflow owners;
- governance and safety layers;

- marketplaces for agents, tools or services.

They know this. Their strategic behaviour already shows it.

The race is not just to build a better model.

It is to avoid becoming a commodity supplier of intelligence.

But that is hard, because the companies best positioned to own distribution and context are often not the model labs. They are Apple, Microsoft, Google, Amazon, Meta, Samsung, device makers, enterprise software incumbents, cloud platforms and eventually AI-native operating companies.

The model lab starts with magic.

The platform company starts with the customer.

In mature markets, the customer relationship is usually the better place to stand.

## **The API story breaks when intelligence becomes ambient**

The current AI economy is still destination-based.

You go to ChatGPT. You go to Claude. You open the app. You call the API. You paste the document. You ask the question.

That feels natural because the technology is still young.

The mature AI economy will be more ambient.

The intelligence will sit inside the editor, the browser, the operating system, the CRM, the ERP, the calendar, the inbox, the file system, the call centre, the warehouse, the IDE, the design tool, the finance process, the compliance workflow and the business control plane.

When that happens, the user does not experience AI as a model endpoint.

They experience it as a capability of the environment.

That is the point at which the API moat weakens.

The question stops being:

*Which model is best?*

And becomes:

*Which system has the data, permissions, workflow, user trust and execution rights to act?*

That is a much harder question for pure model labs.

## **Intelligence becomes a component**

The final form of AI may not be a product category called AI.

It may be a component in almost every product category.

That is what happens when technologies mature. Electricity disappeared into devices. Networking disappeared into software. Databases disappeared into applications. Cloud disappeared into product delivery. Machine learning disappeared into recommendations, fraud detection, ranking, search and logistics.

AI may disappear too.

Not because it becomes less important.

Because it becomes assumed.

That is the danger for companies whose valuation depends on intelligence remaining visible, scarce and separately priced.

In the end, the moat may not be the model.

It may be the place where the model runs, the data it can see, the actions it is allowed to take, the workflow it belongs to and the trust structure around it.

That is why I think intelligence as an API is a transitional phase.

It is powerful.

It is necessary.

It is not the final ownership layer.

In [Part 4](#), I follow that logic into enterprise software and ask what happens when software itself becomes easier to generate around the business.

## PART 4

# SaaS Was a Compromise

---

## Part 4: SaaS Was a Compromise

[Part 3](#) argued that the durable moat in AI is unlikely to sit neatly inside the model API. This chapter shifts the lens to application software and asks what happens when the economics that made SaaS dominant begin to change, before the company-form consequences become explicit in [Part 5](#).

SaaS did not win because generic software was perfect. SaaS won because custom software was too expensive.

That distinction matters because AI attacks the second part of the sentence.

For the last two decades, businesses have increasingly adapted themselves to software. A company buys a CRM, an ERP, a ticketing system, a project management tool, a marketing automation platform, a support platform, a finance system, a workflow tool, a reporting tool, a vertical SaaS product, and then the business bends around the assumptions inside those products.

- The software defines the workflow.
- The company configures what it can.
- The rest becomes process change.

This was not irrational. It was the best available compromise. Building custom software was slow, expensive, risky and difficult to maintain. Buying SaaS meant accepting a generic model in exchange for speed, reliability, upgrades, integrations, compliance, support and predictable subscription pricing.

That bargain created enormous companies, but the bargain depends on one assumption:

*It is cheaper to adapt the business to the software than to adapt the software to the business.*

AI threatens that assumption.

## The old software compromise

Every company has a slightly different shape.

Different customers. Different products. Different sales processes. Different risk models. Different pricing logic. Different compliance requirements. Different internal language. Different approval paths. Different exceptions. Different legacy systems. Different reporting needs. Different politics.

But software markets prefer repeatability.

A SaaS company cannot build a completely bespoke product for every customer and still enjoy SaaS economics. It needs a product model that is reusable across many customers. It needs standard objects, standard workflows, standard configuration surfaces, standard permissions, standard integrations and standard pricing.

That creates a mismatch. The business is specific. The software is generic.

The gap is filled by configuration, implementation partners, consultants, spreadsheets, manual workarounds, exported CSV files, integrations, process changes and people whose job is to make the tool fit reality.

This is the hidden tax of SaaS. We often talk about subscription cost, but the real cost is broader:

- implementation;
- migration;
- training;
- process redesign;
- integration;
- administration;
- customization;
- reporting workarounds;
- manual exception handling;
- vendor lock-in;
- duplicated data;

- the mental overhead of living inside someone else's model of your business.

SaaS is useful because the alternative was worse.

AI changes the alternative.

Layer	SaaS-era default	AI-native direction
Software design	Buy a generic product and configure it	Generate or compose software closer to the operating model
Workflow fit	The business bends around the tool	The tool bends back toward the business
Economic logic	Reuse one application across many customers	Lower the cost of bespoke logic, integration and maintenance
Hidden tax	Implementation, workarounds, admin, duplicated effort	Governance, evaluation, orchestration and quality control
Scarce asset	The packaged workflow	Trusted records, permissions, policy and execution rights

## FIGURE

## The buy-versus-build threshold

Three threshold zones determined by workflow specificity and the cost of custom software. As AI reduces the cost of building, high-specificity workflows shift toward the build side.

### 1 Buy the packaged workflow

#### WORKFLOW IS GENERIC · CUSTOM SOFTWARE STILL EXPENSIVE

The workflow is generic enough, and custom software is still expensive enough, that SaaS remains the rational default.

**IMPLICATION** · This is the world that made the old SaaS compromise so strong.

### 2 Keep the record, generate the workflow

#### SPECIFICITY RISING · CUSTOM SOFTWARE COST FALLING

The core system of record may still be bought, but the workflow layer starts moving toward generated internal capability.

**IMPLICATION** · This is the transition zone where generic workflow SaaS gets pressured first.

### 3 Build or generate around the business

#### HIGH SPECIFICITY · CUSTOM SOFTWARE CHEAP

The workflow is specific enough, and custom software cheap enough, that the economics start favouring software shaped around the company itself.

**IMPLICATION** · This is where the SaaS workaround loses scarcity.

#### What the old SaaS era assumes

It is cheaper to adapt the business to the software than to adapt the software to the business.

#### What AI changes

When workflow specificity stays high while custom software costs keep falling, the workflow layer becomes much easier to generate around the company itself.

## Software starts adapting to the business

The most important AI shift in software is not that a chatbot appears inside every application.

It is that software itself becomes easier to create, change and maintain.

That does not mean a non-technical person can safely prompt a complete enterprise system into existence and walk away. That is fantasy.

But it does mean the cost curve changes.

AI can accelerate requirements capture, prototyping, code generation, testing, documentation, migration, integration, refactoring, support, analysis and maintenance. It can help translate messy business language into working software. It can generate UI, APIs, database schemas, workflow logic, test cases, documentation and deployment scripts. It can inspect existing systems and suggest adapters. It can create internal tools that previously would have been uneconomical.

When the cost of custom software falls, the boundary between buy and build moves.

That is the threat to SaaS. Not that every company suddenly becomes a software company in the old sense. Rather, every company gains the ability to express more of its own operating model directly in software.

The software starts adapting to the business. That is a reversal of the SaaS era.

## **Systems of record versus systems of workflow**

Not all SaaS is equally exposed.

Some systems are deeply embedded, regulated, audited and structurally hard to replace. Finance cores, ledgers, payroll, tax systems, HR master data, identity systems, banking infrastructure, ERP cores and compliance-heavy systems of record have a different durability.

These systems do not survive merely because they have nice interfaces. They survive because they become the legal, financial or operational memory of the company.

Replacing them is not just a software decision. It is a governance decision, a compliance decision, a migration risk and an operational risk.

But systems of workflow are more exposed.

A large part of the SaaS market is not a system of record in the deep sense. It is workflow around records.

Forms. Dashboards. Approvals. Notifications. Ticket queues. Lightweight CRMs. Internal portals. Reporting layers. Project boards. Operational trackers. Lead routing. Customer onboarding flows. Support triage. Data enrichment flows. Content calendars. Review processes. Procurement workflows. Risk review workflows. Contract handoffs.

Much of this is structured CRUD plus workflow plus permissions plus reporting plus integrations.

That pattern used to be expensive to build well. AI makes it more generatable. That does not make it trivial. It does make it vulnerable.

## **The SaaS feature becomes the AI-generated workflow**

A SaaS company normally packages a workflow and sells it repeatedly.

AI makes it easier for a company to generate a workflow around its own exact requirements.

Imagine a company that wants a customer intake process.

In the SaaS world, it buys a product, configures fields, creates workflows, trains staff, integrates with existing systems and accepts the tool's assumptions.

In an AI-native world, the company describes the process, feeds in examples, policies, documents, APIs, data models and exceptions, then generates a workflow that matches its operating model.

The system creates the intake UI, validation logic, enrichment calls, approval flow, audit trail, customer communication, reporting dashboard and integration events.

A human still reviews. Engineers still govern. Security still matters. Compliance still matters. Testing still matters. But the economics are different.

What used to be a SaaS feature can become a generated internal capability.

That is the phrase that should worry SaaS companies:

*generated internal capability.*

Because that is where SaaS loses scarcity.

## Seat-based pricing breaks under labour compression

SaaS economics have historically loved headcount.

More employees means more seats. More seats means more recurring revenue. Expansion revenue becomes a growth engine. Net revenue retention becomes a valuation story. The software company grows as the customer grows.

AI complicates that.

If AI reduces the number of humans required to perform a workflow, seat-based pricing comes under pressure.

A customer using AI to handle support triage may need fewer support agents. A customer using AI to automate sales operations may need fewer operations seats. A customer using AI to generate reports may need fewer analysts. A customer using AI to draft, classify, review and route work may need fewer licensed users touching the system.

The SaaS vendor will try to respond with usage pricing, outcome pricing or AI add-on pricing.

But then the customer asks a harder question:

*If the AI is doing the work, why am I paying per human seat?*

And then the next question:

*If the workflow is now mostly automated, why am I buying a generic workflow product at all?*

That is the structural problem.

AI is not just an upsell opportunity for SaaS.

It is also a force that attacks the pricing model underneath SaaS.

## AI features are not moats

Every SaaS company now has an AI strategy.

### **That is not a moat.**

In many cases, it is table stakes. Adding summarization, drafting, chat, classification, recommendations or natural language search to an existing product may improve the product. It may defend the product for a while. It may create short-term pricing power. It may keep customers from leaving.

But if every product has AI, then AI is not the differentiator.

The differentiators become:

- proprietary data;
- deep workflow ownership;
- system-of-record status;
- trust;
- compliance;
- integration depth;
- permissions;
- governance;
- ecosystem;
- distribution;
- switching costs;
- ability to execute real actions safely.

The weak SaaS companies are the ones whose real product is a generic workflow wrapped around data they do not uniquely own.

AI can replicate that more easily than investors want to admit.

## Vertical SaaS is both strong and vulnerable

Vertical SaaS is interesting because it cuts both ways.

On one hand, vertical SaaS companies can have deep domain knowledge, specialized workflows, regulatory understanding, data models, integrations and customer trust. That gives them defensibility.

On the other hand, many vertical SaaS products are highly specific workflows. That means they are exactly the kind of software AI may eventually generate more cheaply for narrow use cases.

The question is whether the vertical SaaS company owns a durable system of record and network of integrations, or whether it mostly owns a configurable workflow pattern.

If it owns the record, the data, the compliance model, the ecosystem and the customer relationship, it can survive and embed AI.

If it owns a workflow that customers tolerate because custom software was too expensive, it is vulnerable.

This is why the SaaS market will not collapse evenly.

It will **sort**.

Systems of record get defended. Systems of workflow get compressed.

Thin workflow SaaS gets replaced, absorbed or forced downmarket.

## **The enterprise control plane becomes more important**

As software becomes more generatable, governance becomes more important.

This may sound counterintuitive, but it is crucial.

If every department can generate workflows, tools, dashboards, automations and agents, the company needs a way to control what is created, what data it can access, what actions it can take, how it is audited, how it is deployed, how it is monitored and how it is retired.

The future is not random prompt-built shadow IT everywhere.

That would be chaos.

The future is an enterprise control plane: identity, permissions, policy, data access, workflow orchestration, observability, audit, evaluation, deployment, rollback, human approval gates, tool catalogs, API governance and model routing.

This is where the value may move. Not necessarily to the SaaS application itself, but to the layer that lets a company safely generate, run and govern its own software and AI workflows.

That is why the SaaS threat is not simply replacement by AI.

It is replacement by an AI-native operating model.

## **SaaS becomes less of an application category and more of a component supplier**

In the mature AI economy, SaaS may not disappear, but it may become more modular.

A company might still use Salesforce as a customer record, Workday as an HR record, SAP as a finance and ERP core, ServiceNow as an IT service backbone, Microsoft as productivity and identity infrastructure.

But around those systems, the workflow layer becomes more fluid. Instead of buying another SaaS product for every operational variation, companies generate workflows that read from and write to systems of record through governed APIs.

The SaaS product becomes a component in a larger AI-operated environment. That is a smaller and less glamorous role than many SaaS valuations imply.

It is still valuable, but it is not the same as owning the user's entire workflow.

## **The custom software renaissance**

For years, buying SaaS was treated as modern and building custom software was treated as dangerous, expensive and old-fashioned.

That made sense in context, but AI may trigger a custom software renaissance.

Not the old custom software model of huge teams, multi-year projects, brittle codebases and expensive consultants.

A new model appears instead: generated scaffolding, AI-assisted engineering, reusable components, workflow engines, API-first systems, automated tests, infrastructure templates, continuous refactoring, observability from the start, human review and governance, and smaller teams maintaining more software.

This does not eliminate engineering. It changes the leverage of engineering.

A company with strong internal technical capability and clear domain knowledge may be able to build exactly what it needs at a cost that would have been impossible before.

That is bad for generic SaaS. It is good for companies that understand their own business deeply.

## The end of the generic workflow premium

The core SaaS risk is not that software stops mattering.

Software matters more.

The risk is that generic workflow software loses its premium because software becomes easier to shape.

That is the **inversion**.

The SaaS era said:

*Buy the best generic workflow and adapt your company around it.*

The AI-native era says:

*Describe your company and generate the workflow around it.*

Not everywhere. Not instantly. Not without governance. Not without engineers, but enough to change the market.

This is why SaaS is more vulnerable to AI than many people believe.

The threat is not merely that AI features appear inside SaaS products. The threat is that AI changes the economics that made SaaS the default choice.

SaaS was a compromise for the high cost of custom software.

If that cost falls, the compromise gets repriced.

In [Part 5](#), I take that repricing logic one step further and ask what happens when the company itself starts becoming software.

## PART 5

# Firmware

---

## Part 5: Firmware

[Part 4](#) argued that AI changes the economics of software. This chapter pushes the argument into organizational form: if software becomes easier to generate and workflows become easier to automate, the company itself starts changing shape. That is the bridge into the timing question of [Part 6](#).

Every company runs on firmware.

Not the kind you update on a router. The kind embedded in how the company actually operates: the processes that run without being written down, the coordination patterns that exist before they appear in any org chart, the institutional logic that tells everyone what to do when something unexpected happens.

Call it firm-ware: the operating logic embedded into the company itself.

Historically, that logic was carried by people. Managers, coordinators, administrators, specialists. Each one a node in a human network that turned goals into repeated action.

AI makes firm-ware programmable. That is the step people still underestimate.

The next great AI startup may not be a company that sells AI. It may be a company that is mostly run by AI.

Most of the AI conversation is about tools inside existing companies. AI for customer support. AI for developers. AI for sales teams. AI for analysts. AI for marketers. AI inside CRM. AI inside office software. AI inside design tools. AI inside search. AI inside every SaaS product that can add a chat box and a sparkle icon.

That is the first-order effect. The second-order effect is more important.

If AI can create software, operate workflows, coordinate tasks, monitor markets, communicate with customers, generate reports, analyze data, handle exceptions, run experiments and improve processes, then the structure of the company itself changes.

A company is, among other things, a coordination machine. AI attacks coordination.

## **The company as a control system**

A business is not just a collection of workers. It is a system for turning goals into repeated action.

A company finds customers, sells to them, serves them, bills them, supports them, reports on them, learns from them, hires people, manages cash, handles risk, negotiates contracts, builds tools, improves processes and tries to survive long enough to do it again tomorrow.

Historically, that required people at almost every layer. Some people did the work. Other people coordinated the work. Other people reported on the work. Other people managed the people coordinating the work. Other people bought software to track the work. Other people integrated that software. Other people produced decks explaining why the work was behind schedule.

The coordination tax is enormous.

It hides inside meetings, planning sessions, tickets, status updates, dashboards, reporting packs, approval chains, handovers, process documents, onboarding, training, vendor management and management layers.

AI does not only reduce task labour. It reduces the coordination tax. That is why the AI-operated company matters.

## **The serious version of the DAO idea**

Crypto had an early version of this idea with DAOs.

A DAO was supposed to be a decentralized autonomous organization. It had treasury, voting, governance proposals, smart contracts and programmatic execution.

The idea was interesting but the execution was often primitive.

Most DAOs did not have operational intelligence. They had governance mechanics without a brain. They could vote, allocate funds, execute simple contracts and coordinate communities, but they could not truly run a business. They could not reliably sell, support, build, negotiate, learn, adapt, inspect markets, generate software or manage complex operations.

AI supplies the missing layer.

An AI-operated company does not need to be a crypto DAO. It may not need a token. It may not need decentralization. It may be a normal GmbH, Ltd or LLC with a legal wrapper, bank account, directors, accountants, contracts and liabilities.

- The difference is internal.
- The operating core is automated.
- Humans set goals, constraints and approvals.
- AI executes much of the company.

That is a more plausible version of the autonomous organization than most of what crypto produced.

## **The sequence: from code to company**

The path from coding assistant to AI-operated company is not magic.

It is a sequence.

- First, AI helps write software.
- Then AI helps maintain software.
- Then AI helps integrate systems.
- Then AI helps generate internal tools.
- Then AI starts operating workflows inside those tools.
- Then AI monitors outcomes.
- Then AI suggests process changes.
- Then AI implements some of those changes.
- Then AI coordinates between functions.
- Then AI handles more of the customer lifecycle.
- Then AI runs experiments, evaluates results and adjusts operations.

At each stage, humans remain involved, but the ratio changes.

The company becomes less like a hierarchy of people using software and more like a set of goals, rules, workflows, data, models, tools and human approval gates.

That is a different company form.

## Minimal human capital

I do not think the realistic version is zero humans. That is clearly a distraction.

There will be humans because law, trust, accountability, taste, relationship, judgement, liability and capital allocation still matter.

But the number of humans required to operate a meaningful business may fall dramatically.

- A five-person company may behave like a fifty-person company.
- A ten-person company may behave like a two-hundred-person company.

A solo founder may operate a portfolio of small AI-run businesses that would previously have required teams.

Human roles shift upward:

- founder;
- goal setter;
- domain expert;
- capital allocator;
- brand and trust owner;
- compliance owner;
- escalation point;
- customer relationship owner;
- auditor;
- final approver for high-risk decisions.

The work does not disappear entirely.

The leverage changes.

That leverage is economically explosive.

<b>Company layer</b>	<b>Human-heavy default</b>	<b>AI-operated direction</b>
Execution	Many people performing repeated tasks	More workflows executed by software and agents

<b>Company layer</b>	<b>Human-heavy default</b>	<b>AI-operated direction</b>
Coordination	Meetings, tickets, handoffs, management layers	Policy, routing, monitoring and approval gates
Software surface	Large SaaS stack plus manual glue	Operating core plus generated internal interfaces
Human role	Do the work, track the work, explain the work	Set goals, approve risk, own trust, handle escalation
Economic effect	Headcount and coordination tax compound	Small teams can operate much larger surfaces

---

**VISUAL GUIDE**

## From human-heavy company to AI-operated company

This is not a literal org chart. It is a leverage map showing which layers shrink, which layers become software, and where the remaining human roles concentrate.

---

**HUMAN-HEAVY DEFAULT**
**Execution surface**

Many people perform repeated operational tasks inside a growing SaaS stack.

**Coordination layer**

Meetings, tickets, reporting packs, and management layers absorb a large coordination tax.

**Human role concentration**

Humans both do the work and supervise the work, so headcount scales with output more quickly.

---

**AI-OPERATED DIRECTION**
**Operating core**

Workflow engine, model routing, observability, approvals, and generated internal interfaces replace large parts of the old stack.

**Human concentration**

Humans move upward into goal-setting, trust, domain judgment, risk ownership, and final approval.

**Economic result**

Small teams can operate larger business surfaces, compressing both labour cost and coordination cost at the same time.

---

**What shrinks**

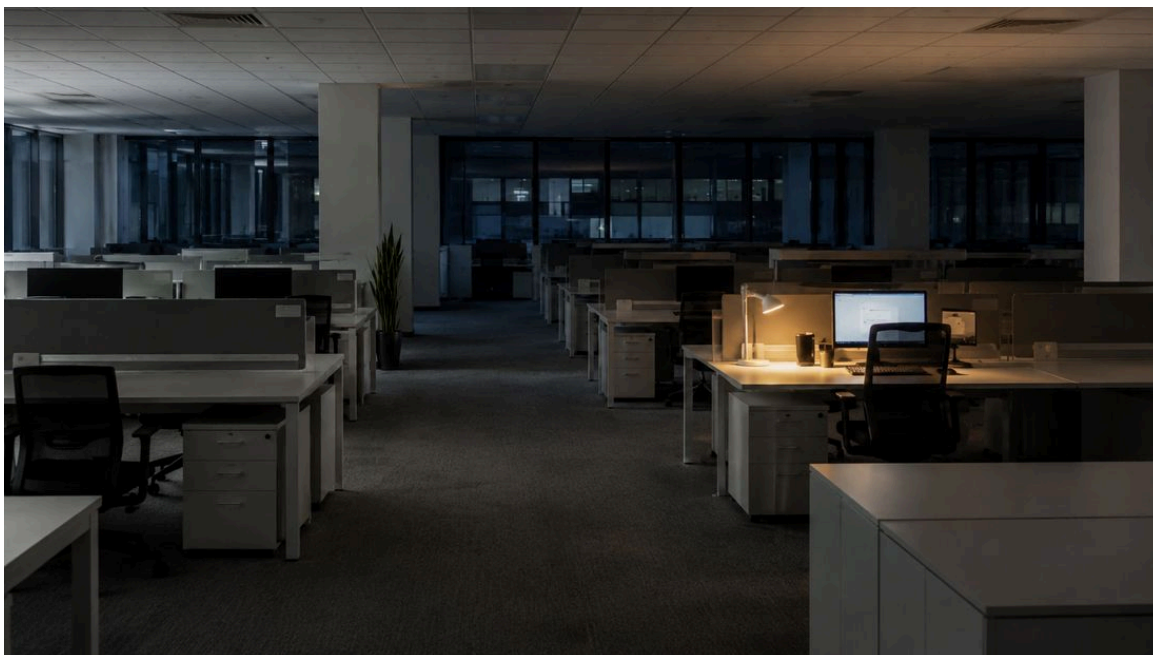
Repeated task labour, reporting overhead, and coordination-heavy middle layers.

**What grows**

Software-defined execution, policy, routing, monitoring, and approval systems.

**What stays human**

Accountability, taste, trust, capital allocation, legal responsibility, and high-risk judgment.



*The AI-operated company is not only a different cost structure. It has a different atmosphere of coordination.*

## **Early categories**

The first AI-operated companies will not be nuclear plants, airlines, hospitals or banks.

They will start where the product is digital, the work is measurable and the liability is manageable.

Likely early categories include:

- niche research businesses;
- market monitoring services;
- regulatory tracking;
- lead enrichment;
- supplier intelligence;
- customer intelligence;
- small software products;
- generated internal tools;
- content operations;
- localization;
- analytics services;

- support operations;
- sales operations;
- compliance screening;
- document processing;
- API integration services;
- code migration;
- dependency updating;
- testing and documentation services.

Many of these are currently human-heavy services or small SaaS categories.

That makes them exposed.

An AI-operated competitor does not need to look impressive from the outside. It only needs to deliver the service with a lower cost structure, faster iteration and fewer humans.

That is enough to change a market.

## **The agency model gets compressed**

Agencies and consultancies are especially interesting.

A lot of agency work is coordination, production and iteration. Content calendars, campaign drafts, reports, SEO pages, analytics, outreach, design variants, customer research, competitor monitoring, dashboarding, workflow automation, integration projects, documentation and project management.

Historically, an agency sold access to people who could do this work.

An AI-operated agency sells outcomes produced by a small human team and a large automated execution surface.

That is a different margin structure.

- It can undercut traditional agencies.
- It can serve smaller niches.
- It can run continuously.
- It can test more variants.

- It can personalize more deeply.
- It can turn expertise into repeatable workflows more quickly.

This does not mean every agency disappears. Trust, taste, strategy, relationships and high-end creative judgement still matter.

But the middle of the market compresses.

The same pattern will apply to many human-heavy service businesses.

## **The SaaS stack gets replaced by an operating core**

A normal small or medium business accumulates SaaS.

CRM, email, project management, support desk, analytics, accounting, workflow tools, automation tools, document management, marketing tools, sales tools, dashboards and internal portals.

The stack grows because every operational need maps to another product category.

An AI-operated company may look different.

It may have:

- a system of record;
- a data store;
- identity and permissions;
- a workflow engine;
- a tool and API catalog;
- model routing;
- observability;
- audit logs;
- generated internal interfaces;
- human approval gates;
- customer-facing surfaces generated as needed.

The company does not buy a separate SaaS product for every workflow. It generates and operates workflows around its own goals.

That is why the AI-operated company connects directly to the SaaS argument. If the company can generate its own operating machinery, the SaaS stack becomes less central.

## **The startup factory becomes real**

The AI-operated company also changes company formation.

A founder may be able to test more ideas with less capital. A venture studio may be able to launch more experiments with fewer employees. A corporate innovation team may be able to generate and test internal products more quickly. A holding company may be able to operate many small businesses with a shared AI operating layer.

This is not the same as the old dream of no-code.

No-code allowed people to build applications without traditional programming, but it often hit limits around complexity, integration, maintainability and customization.

AI-assisted software and workflow generation goes further because it can operate across code, data, text, APIs, tests, documentation and process.

The bottleneck moves from building the first version to governing the system.

The winners will be people and companies that can define clear goals, constraints, quality bars, data boundaries, approval rules and operating metrics.

The scarce skill becomes not just coding.

It becomes knowing what should exist.

## **Labour is the shadow of this argument**

This series is mainly about capital markets and business models.

But it is impossible to discuss AI-operated companies without seeing the labour implications.

If companies can produce more output with fewer people, employment structures change. If coordination layers shrink, management changes. If AI can perform knowledge work tasks at scale, entry-level career ladders change. If small teams can operate company-sized surfaces, competition changes. And if AI-operated companies can undercut human-heavy companies, wage pressure appears in affected sectors.

This deserves its own essay, probably its own series. But it sits behind everything.

The AI bubble is not just a stock market question. It is a question about the cost structure of the firm.

When the cost structure of the firm changes, capital markets eventually notice.

## **Existing companies are not safe**

It is tempting to assume incumbents win because they have customers, data, brand and distribution.

Some will.

But many incumbents also have bureaucracy, legacy systems, internal politics, heavy cost structures, slow decision-making, risk aversion, vendor lock-in and processes designed for a pre-AI world.

An AI-operated competitor can be structurally different. It can start with a smaller team, a cleaner stack, generated software, automated workflows, lower overhead and a more adaptive operating model.

This does not guarantee victory, but it creates pressure.

The internet did not simply give existing retailers a new channel. It also created Amazon. AI will not simply give existing companies better tools. It will create new company forms.

That is where the deeper disruption sits.

## **The company becomes software**

For years people have said every company is a software company, but that was only partly true.

Most companies used software. Some built software. Very few became software in a deep operational sense.

AI may make the phrase more literal. A company can become a set of encoded goals, workflows, policies, data pipelines, models, agents, approval gates, observability systems and generated interfaces.

It can inspect itself, modify parts of itself, run experiments on itself, generate tools for itself and coordinate work through software rather than through layers of meetings and management.

This is not utopian. It may be messy, dangerous and highly unequal. It will create new risks around opacity, accountability, dependency, labour and control.

But as a business form, it is powerful. The market is currently focused on who sells AI.

The larger question is who uses AI to become a different kind of company.

## Agent orchestration makes the company form real

One concrete example of the infrastructure I am describing is [Paperclip](#), an open-source control plane that treats agents as employees. Paperclip provides org charts, budgets, heartbeats, atomic execution, persistent sessions and audit logs so you can "hire" agents, assign goals and govern work rather than babysit tabs. In that world the emerging class of claw-style adapters — lightweight, webhook- or runtime-driven agent connectors such as [OpenClaw](#) — act less like demos and more like modular workers you can plug into an operating system for the company.

That practical difference matters. These runtimes make agents durable. Scheduled heartbeats, checked-out tasks, resumable session state and cost controls turn ephemeral model sessions into repeatable work products. They also surface the real governance questions: who audits agent decisions, who owns budgets, how do you rollback bad changes, and what layer keeps the whole system observable enough to trust?

The market is already reacting to that shift. In February 2026 the creator of OpenClaw, Peter Steinberger, announced he was joining OpenAI while OpenClaw itself remained open source with independent stewardship, a move widely read as evidence that larger model labs want a deeper foothold in agent runtimes. Anthropic's recent product direction points the same way. The reason is straightforward: once orchestration, governance and durable runtimes become central, the profit pool moves upward. Whoever controls safe, observable, auditable AI labour captures more value than a company that merely sells model access.

The AI-operated company matters to the bubble thesis because it shifts the profit pool.

If AI's most important economic effect is not simply selling model access or adding features to SaaS, but reducing the human and software cost of operating a company, then the current market story is incomplete.

The value does not necessarily accrue to the model API. It does not necessarily accrue to the SaaS vendor. It may accrue to the company that uses AI to operate with a radically lower cost structure.

That is dangerous for both model labs and SaaS incumbents.

Model labs may become suppliers. SaaS vendors may become components. AI-operated companies may capture the customer.

That is the deeper convergence. The technology succeeds, but the old valuation story breaks under the weight of a new company form.

In [Part 6](#), I map the pressures that turn that structural mismatch into an actual market break.

## PART 6

# Breaking Point

---

## Part 6: Breaking Point

[Part 5](#) argued that AI changes not only software, but the cost structure of the firm itself. This chapter turns from structure to timing: what actually causes the market story to break. It is the hinge in the series, and the natural bridge into the post-crash world of [Part 7](#).

The market does not stop believing in AI. It starts doubting who owns it.

That is the mechanism behind every belief break in a real-paradigm bubble. Not a sudden rejection of the technology, but a quiet erosion of the story about ownership. And when enough of those stories erode at the same time, the valuation can no longer hold.

That is the point people often miss. Markets can live with overvaluation for a long time if the story still works. Investors can tolerate losses, burn rates, dilution, aggressive forecasts, thin margins and stretched multiples if they believe the current discomfort is the price of owning the future.

The break comes when the story stops explaining reality. Not all at once. Not usually through one single event. More often, several pressures build until the shared belief shifts.

In the dotcom bubble, the internet did not suddenly become less important. The belief that changed was more specific.

Investors stopped believing that every company with a domain name was a future monopoly.

They stopped believing that traffic alone was enough, that customer acquisition cost did not matter, that first mover advantage could overcome bad economics, and that capital could buy time forever.

The future was still real. The investable story broke. That is what I expect with AI.

The AI bubble will not burst because people stop using AI. It will burst when enough people realize that the current market story misidentifies who owns the future.

## The market story today

The current AI story is a bundle of beliefs.

- It says frontier models remain scarce and valuable.
- It says the companies with the best models can capture intelligence rents through APIs and subscriptions.
- It says hyperscaler capex is a sign of future demand rather than a possible overbuild.
- It says the chip and data centre buildout will keep compounding because AI demand is effectively insatiable.
- It says SaaS companies can defend their workflow ownership by adding AI features.
- It says enterprises will move from pilots to deep adoption fast enough to justify the infrastructure.
- It says AI agents will increase usage, increase productivity and remain dependent on centralized cloud intelligence.
- It says the first visible AI winners are probably the long-term owners of AI value.

That is a lot of story.

Each piece can be partially true.

The bubble breaks when too many pieces become less true at the same time.

## Pressure 1: The model scarcity premium declines

Today, frontier models feel scarce.

Only a small number of companies can train and serve the best systems. That scarcity supports the idea that model labs can own the intelligence layer.

But the scarcity premium is under attack.

Models get smaller. Open models improve. Specialized models become good enough for narrow tasks. Distillation moves capability downward. Inference gets cheaper. Hardware improves. On-device and enterprise-local models become more practical.

The frontier keeps moving, but commercial value does not always require the frontier.

Many tasks require reliable, cheap, contextual, private, low-latency intelligence rather than the smartest possible model.

As that becomes obvious, the market will ask how much pricing power belongs to the model API.

If the answer is less than expected, the story weakens.



## **Pressure 2: Token economics become visible**

A lot of AI adoption is still happening under soft economics.

Flat-rate subscriptions. Promotional pricing. Cloud credits. Internal experimentation budgets. Strategic subsidies. Bundled features. Investor-funded usage. Enterprise trials. Internal mandates to test AI.

This creates activity, but not always durable economics.

The hard question is what happens when usage is priced honestly.

Agentic workflows can burn tokens. Coding agents can burn tokens. Research agents can burn tokens. Customer support automation can burn tokens. Multi-step planning, retrieval, retries, tool calls, evaluation and monitoring all cost money.

If the output is valuable enough, this is fine.

But if much of the usage is exploratory, low-quality, redundant or requires expensive human correction, then customers will become more disciplined.

The AI market is currently full of enthusiasm about usage.

The convergence point comes when usage has to justify itself as margin.



## **Pressure 3: Enterprise ROI disappoints before it matures**

AI is useful in enterprises.

That does not mean enterprise transformation is easy.

Real companies have messy data, legacy systems, compliance requirements, permissions, audit needs, security constraints, customer promises, regulatory obligations, internal politics and edge cases.

The demo is always cleaner than the business.

A chatbot that summarizes a document is impressive. A governed AI workflow that safely acts across customer data, legal constraints, pricing rules, finance systems, support systems and audit requirements is much harder.

The value is there, but it requires integration. The likely pattern is disappointment first, maturity later.

Many pilots will fail to produce measurable returns. Some internal AI programmes will become expensive experiments. Some consultancies will sell transformation without operational depth. Some companies will discover that the model was not the hard part. The hard part was workflow, data, permissions, governance, evaluation and change management.

This does not kill AI. It slows the revenue curve.

If valuations assume rapid enterprise conversion, that matters.



## **Pressure 4: Hyperscaler capex becomes a question instead of proof**

Right now, AI infrastructure spending is treated as proof of demand.

The logic is simple: Microsoft, Google, Amazon, Meta and others are spending enormous amounts on data centres, chips, networking, power and cooling. They must see demand. Therefore the buildout validates the market.

That is true until it is not. At some point, capex changes meaning.

If utilization disappoints, if customers resist pricing, if model efficiency improves faster than demand, if local inference absorbs common workloads, if financing costs rise, if power constraints bite, or if companies guide toward more discipline, the same capex can be reinterpreted as overbuild.

This is psychologically important.

The market currently sees spending as confidence. The bubble cracks when spending starts to look like sunk cost.

The language will be subtle at first:

- optimizing utilization;
- moderating future expansion;
- focusing on capital discipline;
- aligning capacity with demand;
- improving efficiency;
- prioritizing high-return workloads.

None of those phrases means AI is dead. They mean the growth story is changing.



## **Pressure 5: SaaS growth slows and pricing power weakens**

SaaS is part of the AI bubble even when it is not labelled that way.

Many SaaS companies are priced as if AI will enhance their existing products, defend their workflows and create new revenue streams.

That may be true for the strongest platforms.

But AI also attacks SaaS.

- If AI reduces headcount, seat-based pricing weakens.
- If AI generates custom workflows, generic SaaS workflows lose scarcity.
- If AI features become standard, they stop being premium add-ons.
- If enterprises become more capable of building internal tools, the buy-versus-build boundary shifts.
- If AI-operated companies use thinner software stacks, traditional SaaS expansion weakens.

The pressure may appear first as slower expansion, lower net revenue retention, more pricing scrutiny, consolidation of tools, fewer seats, longer sales cycles or customer resistance to AI surcharges.

That does not look like a dramatic crash at first. It looks like growth becoming ordinary.

For high-multiple SaaS, ordinary growth is dangerous.



## **Pressure 6: Thin AI wrappers fail**

Every paradigm shift creates wrappers.

A wrapper is a company that packages a new technology into a narrow product before the durable platform structure is clear.

Some wrappers become real companies. Many do not.

The AI wrapper problem is obvious. A startup takes a model API, adds a workflow, a prompt library, a UI, a narrow use case, maybe some integrations, then raises capital as if it has created a durable product.

Some of these will work.

Many will be crushed by model providers, SaaS incumbents, open source, platform features or customers building their own versions.

When the wrapper layer starts failing visibly, the market mood changes.

Again, this does not mean AI is fake.

It means that not every use case is a company.



## **Pressure 7: Financing stress appears below the surface**

The AI buildout is capital intensive.

Data centres, chips, power, cooling, memory, networking and long-term capacity commitments require enormous amounts of money.

Some of this is funded directly by highly profitable companies. Some is funded through debt, leases, partnerships, private capital, infrastructure vehicles, vendor financing and complicated commitments.

In boom phases, this all looks rational. Everyone wants exposure to the buildout. Future demand is assumed. Assets look strategic. Financing looks safe because the tenants and counterparties are high quality.

But infrastructure bubbles can break through financing stress before the public narrative changes.

A data centre project delayed by power constraints. A financing structure that depends on aggressive utilization. A private infrastructure vehicle that cannot roll over debt on expected terms. A supplier with customer concentration. A chip order that gets pushed out. A capacity commitment that looks too large.

These are not necessarily headline events at first.

They are stress signals. The AI bubble may not crack first in a chatbot app.

It may crack in the financing of the physical world underneath it.



## **Pressure 8: The belief shifts from future ownership to future exposure**

This is the psychological turn.

During the boom, exposure to AI is good.

- A company spends on AI infrastructure: bullish.
- A SaaS company adds AI features: bullish.
- A startup uses AI in the pitch: bullish.
- A cloud provider signs a huge capacity deal: bullish.
- A company talks about agents: bullish.

At the convergence point, the meaning changes.

AI exposure becomes a question:

- Are you capturing value or just spending money?
- Are you selling durable capability or subsidized usage?
- Are you defending a moat or training customers to expect cheaper software?
- Are you building infrastructure for future demand or overbuilding ahead of it?
- Are you using AI to lower cost or adding another expensive layer?
- Are your AI features differentiated or table stakes?
- Are you the owner of the workflow or a supplier to someone else's workflow?

That is when the story breaks.

The market stops rewarding exposure and starts demanding proof of capture.



## The convergence

- No single pressure has to be fatal.
- The model labs can survive smaller models.
- Hyperscalers can survive capex discipline.
- SaaS companies can survive slower growth.
- Enterprises can survive failed pilots.
- AI wrappers can fail without damaging the whole market.
- Infrastructure financing can absorb some stress.

The danger is convergence.

Several pressures arriving close enough together that the market narrative cannot absorb them:

- model scarcity declines;
- enterprise ROI slows;
- token economics harden;

- capex growth moderates;
- SaaS expansion weakens;
- wrappers fail;
- financing stress appears;
- investors ask who owns the future.

At that point, the market does not need proof that AI failed.

It only needs doubt that the current winners own enough of the future to justify their valuations.



---

#### BELIEF GAUGE

### Full convergence — eight pressures active

The market no longer asks whether AI matters. It asks whether the current winners own enough of the future. The old story stops carrying the valuation. Exposure flips from a badge of confidence into a demand for proof.

8/8

ACTIVE  
PRESSURES

---

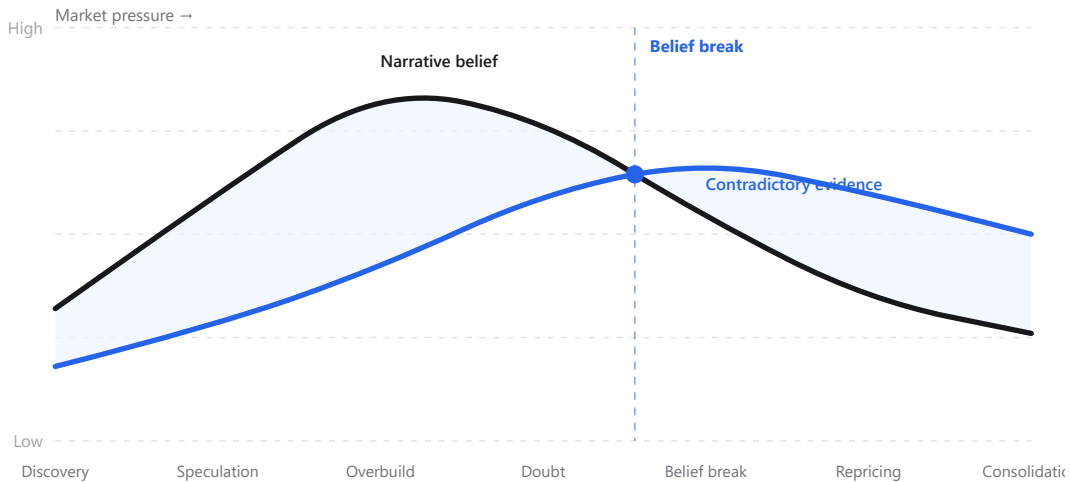
#### INVESTOR QUESTION

Who owns the workflow, the authority, the economics, and the durable control points after the crash?

The chart below is not a financial model. It is a conceptual picture of market psychology. The dark line is narrative belief. The accent line is contradictory evidence. The shaded gap between them is the belief buffer, the market's remaining capacity to explain weak signals away as temporary, strategic, or worth enduring.

**FIGURE****The belief break**

Bubbles survive bad evidence until the story can no longer absorb it. This chart is conceptual rather than numeric. The shaded region between the lines is the belief buffer — when the lines cross, belief breaks and repricing begins.

**Belief buffer**

The shaded gap between the lines. As long as it exists, the market can reinterpret weak signals as temporary noise, transition cost, or the price of owning the future.

**What the market still believes**

The market no longer assumes that AI exposure automatically means future ownership.

**What the evidence is doing**

Contradictory evidence now arrives as a pattern rather than as isolated exceptions.

**What changes at the break**

This is the snap point. The story breaks when the evidence exceeds the market's ability to explain it away. After the break, the question is no longer whether AI matters — it is who survives the repricing with real control.

The chart does not claim precise timing. It visualises the mechanism: contradictory evidence can accumulate for a long time while the story stays intact.

As long as the belief buffer exists, the story survives. Weak ROI can be called early learning. High token burn can be called transition cost. Giant capex can be called confidence. Slower SaaS expansion can be called temporary noise. The break comes when those explanations stop feeling persuasive enough to carry the valuation.

That is how a real-technology bubble bursts.

## The sunk cost problem

Human belief is not rational at the end of a bubble. It is **emotional**.

People have reputations invested. Careers invested. Funds invested. Strategies invested. Public statements invested. Corporate roadmaps invested. Data centre plans invested. Hiring plans invested. Acquisition strategies invested. Product narratives invested.

The more capital is committed, the harder it becomes to admit the story has changed.

This is why bubbles often overshoot and sunk cost becomes identity.

The company cannot easily say the strategy was wrong. The investor cannot easily say the thesis was wrong. The executive cannot easily say the capex was excessive. The founder cannot easily say the moat was temporary.

So the story stretches and every contradiction gets explained away.

Until it cannot be. That is the breaking point.

The moment when maintaining belief becomes more painful than abandoning it.

## The crash will be misread

When the AI bubble bursts, many people will say AI was hype.

That will be the wrong lesson.

The correct lesson will be that the first capital markets story about AI was wrong. The internet after dotcom became more important, not less.

AI after the crash may do the same.

The crash will destroy weak companies, punish bad capital structures, compress valuations, expose fake moats and force discipline, but the technology will continue moving.

Models will improve. Software will become more fluid. AI will become embedded. Companies will automate more. Some workflows will disappear. New business forms will emerge.

The bubble bursts because the story fails. The paradigm continues because the technology is real.

That is the convergence point.

In [Part 7](#), I follow that break forward into the mature AI economy and the assets, power centers and business models that survive it.

PART 7

# After the Crash

---

## Part 7: After the Crash

[Part 6](#) described the convergence that breaks belief. This chapter follows the story through the crash and into the mature AI economy, where the first narrative has failed but the technology keeps spreading. It also sets up the final ownership question answered in [Part 8](#).

The dotcom crash did not end the internet.

It began the mature internet.

That is not how it felt at the time. At the time, it felt like collapse. Companies disappeared. Careers changed. Offices emptied. Investors lost money. Founders who had been treated like prophets became examples of excess. The story broke.

But the infrastructure did not vanish. The habits did not vanish. The technical direction did not vanish. The public did not go back to a pre-internet world.

Instead, the market sobered up.

Weak companies died. Useful assets were absorbed. Valuations reset. Business models became more disciplined. The fantasy layer burned away. The companies that survived had to become real businesses.

Then the internet became more important than ever. That is the part that matters for AI.

The AI crash, when it comes, will not be the end of AI. It will be the beginning of the real AI economy.

## The crash is not the end of the paradigm

A crash feels like rejection. It is often maturation.

In a discovery bubble, the market is not rejecting the technology. It is rejecting the first overfunded story about the technology.

The internet story of the late 1990s had to fail before the mature internet economy could form.

The same may be true of AI.

The first AI story says frontier model labs, hyperscaler data centre buildouts, SaaS incumbents with AI features and thin AI wrappers are the natural owners of the future.

Some of those companies will survive. Some will become stronger. Some will be absorbed. Some will disappear. Some will remain important but less profitable than expected. Some will become suppliers rather than platforms.

The crash will not decide whether AI matters. It will decide which parts of the first story were wrong.

That is a different question.

## **What happened after dotcom**

The internet after dotcom became both more useful and more centralized.

The early web was strange, open, personal, chaotic and independent. It was full of small sites, weird communities, hand-built pages, forums, experiments and businesses that looked nothing like the platform economy that followed.

After the crash came consolidation. Search became a dominant interface. Cloud became the new infrastructure layer. Social networks centralized identity, attention and distribution. Mobile operating systems created new app store gatekeepers.

Then the next layer hardened. Streaming platforms reshaped media. Online advertising became surveillance infrastructure. The consumer became the product. Walled gardens replaced much of the early web's open weirdness.

This is the uncomfortable truth about paradigm shifts. The mature form is often more useful and more extractive than the early form.

The mature internet gave us convenience, scale, search, communication, ecommerce, software distribution, cloud computing, mobile apps, maps, video, payments and access to knowledge.

It also gave us attention markets, behavioural tracking, platform dependency, algorithmic manipulation, creator precarity, app store taxes, surveillance advertising and a handful of companies with extraordinary power.

The crash did not prevent that.

It helped create the conditions for it.

**EDITORIAL FIGURE**

**The investor loses money. The asset survives.**

Bubble capital can disappear while the infrastructure, habits and technical direction remain. The crash clears the fantasy layer. It does not roll the world back.

This is the chapter's core emotional shift: the capital structure can fail even when the technical substrate remains useful.

**INFRASTRUCTURE OUTLIVES THE STORY**

**BUBBLE PHASE VERSUS MATURE PHASE**

Capital rewards exposure to the paradigm	Capital demands proof of durable capture
Capability and narrative dominate	Cost, governance, execution and margins dominate
Infrastructure spend reads as confidence	Infrastructure spend gets judged for timing and utilisation
Many plausible companies get funded	Fewer companies survive, but the technology spreads further
The story is exciting and legible	The power map becomes quieter, more selective and more extractive

The crash does not end the paradigm. It forces the technology into a more selective, more disciplined and often more extractive mature form.

## AI will also centralize after it decentralizes

AI has an interesting tension.

On one side, models will get smaller, cheaper, more local and more embedded. That pushes toward decentralization.

On the other side, the companies that control devices, operating systems, clouds, identity, enterprise systems, data, app stores and workflow platforms are already large. That pushes toward centralization.

Both things can be true.

AI capability may become more distributed while AI power becomes more concentrated.

This is what happened with the internet. Publishing became easier for everyone, but distribution centralized. Communication became easier for everyone, but attention centralized. Software delivery became easier for everyone, but platforms centralized. Commerce became easier for everyone, but marketplaces centralized.

AI may follow the same pattern.

More people and companies will be able to build, automate and create.

But the operating layers may centralize around a small number of powerful control points.

Those control points may include:

- chips;
- device ecosystems;
- operating systems;
- cloud infrastructure;
- app stores;
- identity systems;
- enterprise data platforms;
- workflow orchestration;
- model routing;
- governance and audit;
- payment and billing;

- distribution and discovery.

The mature AI economy may not be owned by the chatbot brands that made AI famous.

It may be owned by the systems that decide where AI runs, what it can see and what it is allowed to do.

## **Weak model labs disappear or are absorbed**

After the crash, the model landscape will likely consolidate.

Training frontier models is expensive. Serving them is expensive. Talent is scarce. Infrastructure is scarce. Distribution is hard. Enterprise trust is hard. Consumer attention is hard. Regulatory compliance is hard. Safety and evaluation are hard.

There will not be unlimited room for every model lab to become a major platform.

- Some will be acquired.
- Some will become specialized providers.
- Some will become research teams inside larger companies.
- Some will pivot to enterprise tooling, safety, evaluation, vertical models, agent infrastructure or deployment platforms.
- Some will disappear.

The surviving labs will need more than model quality.

They will need distribution, trust, cost discipline, enterprise relationships, developer ecosystems, integration depth or control of some strategic layer.

In the mature AI economy, the model is not enough.

## **Infrastructure gets reused**

One of the recurring features of paradigm-shift bubbles is that badly timed infrastructure can still become useful later.

The investor loses money. The asset survives.



*The first story burns off. The physical substrate remains, waiting for the next owner and the next use.*

Dotcom overbuild did not mean the world never needed internet infrastructure. It meant the early capital structure and timing were wrong.

AI data centres, chips, networking, power contracts, cooling capacity and tooling may follow a similar path.

- Some infrastructure may be overbuilt for the first story but valuable for the second.
- Capacity funded by one group may be bought cheaply by another.
- Projects justified by one demand curve may serve a different demand curve later.

This is why the crash should not be confused with technological reversal. The physical and software infrastructure of AI will not vanish.

It will be repriced.

## **Inference becomes cheaper and more invisible**

After the crash, inference costs will matter more.

The fantasy phase celebrates capability. The mature phase optimizes cost, latency, reliability, privacy and integration.

Models will be routed by task. Cheap local models will handle simple work. Specialized models will handle domain-specific work. Frontier models will handle hard cases. Enterprises will build policies that decide what data can go where. Devices will perform more inference locally. Operating systems will bundle AI into default experiences.

The user will see less of the model. AI will become a feature of the environment. That is when it becomes truly powerful.

The most important technologies often disappear into the background.

Electricity did. Networking did. Databases did. Cloud did. Machine learning already did in many areas. AI will too. The bubble version of AI is visible.

The mature version will be ambient.

## **SaaS consolidates and hollows out**

The SaaS market after the AI crash will not simply die.

It will split.

Strong systems of record will remain important. Finance, HR, identity, ERP, compliance, regulated data, core operational records and deep enterprise platforms will still matter.

But workflow SaaS will be under pressure.

Generic workflows become less defensible when custom workflows are easier to generate. Seat-based pricing weakens when AI reduces human participation. AI features become standard. Customers become more willing to ask whether they need another SaaS product or whether they can generate the workflow themselves around existing systems of record.

- Some SaaS companies will consolidate.
- Some will be acquired by larger platforms.
- Some will become data and system-of-record providers.
- Some will become orchestration and governance layers.
- Some will become feature sets inside broader platforms.
- Some will be replaced by generated internal software.

This is not a collapse of software. It is a collapse of the generic workflow premium.

Software becomes more important. Packaged workflow software becomes less automatically valuable.

## **The rise of governed internal software**

After the crash, companies will become more sober about AI.

- They will not stop using it.
- They will use it with more discipline.

That means more governance, not less.

The next phase will be about safe execution:

- which models can be used;
- which data they can access;
- which actions they can take;
- which workflows they can modify;
- which outputs require human approval;
- how performance is measured;
- how failures are audited;
- how generated code is tested;
- how AI tools are deployed and retired;
- how business rules are enforced.

This creates demand for enterprise AI control planes.

Not just chatbots, but control planes.

The companies that can govern AI-generated software, AI-operated workflows and AI agents inside real businesses will be strategically important.

That layer may become one of the mature AI economy's durable profit pools.

## **AI-operated companies become real competitors**

The crash will also make the AI-operated company more plausible.

During the bubble, many people will sell AI as a feature. After the crash, the best operators will use AI as a cost structure.

That is different.

A company with a small human team and a large automated operating surface can compete differently. It can personalize more, iterate faster, serve smaller markets, operate continuously and survive on lower margins.

This will begin in digital services, research, monitoring, content, software, operations, support, data enrichment and narrow B2B services.

Then it will spread.

The interesting thing is that these companies may not advertise themselves as AI companies.

They may simply be cheaper, faster and more adaptive.

That is when the transformation becomes hard to see from the outside but impossible to ignore in the numbers.

## **The consumer becomes more than the product**

The internet taught us that free or cheap services can hide a deeper exchange.

The consumer became the product through advertising, attention, data and behavioural targeting.

AI may deepen that pattern.

The user is not only a target for ads. The user is a source of training signal, behavioural feedback, preference data, workflow data, voice, style, context, intent, correction and evaluation.

- Every interaction can improve the system.
- Every workflow can teach the system.
- Every correction can become part of the model or the product layer around the model.

In the mature AI economy, exploitation may not only mean showing better ads.

It may mean capturing the patterns of how people think, work, decide, write, negotiate, design, code, manage and communicate.

That is a deeper form of extraction. This is where the social consequences begin to overwhelm the capital markets story.

I do not want this series to become primarily about social impact. That deserves its own treatment, but it has to be acknowledged.

The mature AI economy may be useful, profitable and deeply extractive. The dotcom precedent suggests those things can coexist.

## **The crash creates the mature winners**

Before the crash, everyone claims to be building the future. After the crash, the future becomes selective.

Capital becomes disciplined. Customers become demanding. Infrastructure gets repriced. Weak companies lose access to funding. Strong companies buy assets cheaply. Business models have to work. The fantasy layer fades.

That is when the mature winners emerge. They may not look like the bubble winners.

They may be:

- chip and memory suppliers;
- device and operating system owners;
- enterprise identity providers;
- systems of record;
- AI governance platforms;
- workflow orchestration systems;
- companies with proprietary data;
- companies that use AI to operate with radically lower cost;
- platforms that own distribution and execution rights.

Some current winners will remain. Others will be repriced. New ones will appear.

The crash clarifies the map.

## **The mature AI economy will be boring and enormous**

The most important phase of a technology often begins when it becomes boring.

The internet became more powerful when it stopped being a special place you went to and became the substrate of everyday life. Cloud became more powerful when it stopped being a novelty and became the default way software was delivered.

Mobile became more powerful when it stopped being a category and became the normal interface to the world. AI will become more powerful when people stop talking about AI as a separate thing.

It will be inside the software. Inside the device. Inside the workflow. Inside the company. Inside the product. Inside the support process. Inside the operating model.

That is the mature phase. The crash does not prevent it. The crash accelerates it by killing the first story.

That is why I keep returning to the same point. The AI bubble will burst because AI is real.

And after it bursts, the real AI economy begins.

The crash is not where the argument ends. It is where the underlying map becomes easier to see.

In [Part 8](#), I pull the whole series together and map the likely winners, pressured groups and control points in that mature economy.

## PART 8

# The New Power Map

---

## Part 8: The New Power Map

This final chapter gathers the whole series into one question: after the bubble breaks, which layers actually own the durable value? If you are arriving here first, the cleanest entry point is still the [preface](#), followed by [Part 1](#).

The market is looking for the next software company, but it should be looking for the next form of the company itself.

That is the deepest point of this series.

The AI bubble is not just a question of whether Nvidia is overvalued, whether OpenAI can justify its infrastructure plans, whether Anthropic can grow into its valuation, whether Microsoft can monetize Copilot, whether Salesforce can defend CRM, or whether every SaaS company can add enough AI features to keep customers paying more.

Those questions matter.

But the bigger question is where power moves when intelligence becomes cheaper, software becomes more fluid and organizations can operate with less human coordination.

The answer may not be the obvious one.

The mature AI economy may not be owned primarily by the companies that made AI famous.

It may be owned by the companies that control compute, devices, identity, data, governance, workflow and execution.

## The old map

The current map is still shaped by the software and cloud era.

- At the bottom are chips, data centres, networking, storage and cloud infrastructure.
- Above that are model providers and platform services.

- Above that are SaaS applications.
- Above that are business workflows, users and customers.

In this map, value is often captured by packaged software and cloud platforms. Companies buy applications. Employees use seats. Workflows live inside SaaS products. Data lives in systems of record. Integration connects the stack. The model provider appears as a new intelligence layer inside or beside the existing stack.

This is why the first AI story is so tempting.

Just add AI to the map.

Add models to the cloud. Add copilots to SaaS. Add agents to workflows. Add usage pricing. Add more data centres. Add more chips. Add more automation.

But paradigm shifts do not merely add a layer.

**They redraw the map.**

## EDITORIAL FIGURE

**Paradigm shifts do not add a layer. They redraw the stack.**

The first AI story simply inserts model providers into the old software and cloud map. The mature AI economy rearranges the layers around authority, context, execution and control.

**OLD MAP**

Software and cloud-era logic, with AI treated as one more layer to bolt onto the existing stack.

**Cloud infrastructure****Model provider layer****Packaged SaaS applications****Business workflows inside products****Users and customers****NEW MAP**

AI-native logic, where intelligence becomes valuable only when it can run inside trusted systems with permission to act.

**Compute, memory, power and networks****Devices and operating systems****Frontier, local, open and specialised models****Identity, permissions and governance****Systems of record and trusted context****Orchestration, evaluation and execution****Generated workflows and AI-operated companies****What the first story assumes**

If the models are powerful enough, the old cloud and SaaS map can stay basically intact.

**What the mature map rewards**

Control over context, permissions, execution and trusted operating surfaces, not just access to intelligence in the abstract.

**The new map**

The AI-native map looks different.

- At the bottom are still chips, memory, power, networking and data centres. Physical compute matters. Energy matters. Supply chains matter.

- Above that come devices and operating systems, because much intelligence moves closer to the user, the context and the workflow. Then come models, plural: frontier models, local models, open models, specialized models, private models and embedded models. Then come identity, permissions, policy and governance, because intelligence without control is operationally dangerous.
- Above those sit data and systems of record, because AI is only useful when it can see the right context and act on trusted information. Then come orchestration layers, durable workflows, tool catalogs, API gateways, observability, evaluation and human approval gates. Then come generated workflows and generated software. Then come customer-facing services, internal operations and AI-operated companies.

In this map, the model is important, but it is not the whole system.

The durable power sits where **intelligence** meets **authority**.

- Who can see the data?
- Who can decide which model runs?
- Who can grant permission?
- Who can take action?
- Who can observe the result?
- Who can audit the decision?
- Who owns the customer relationship?
- Who owns the workflow?

Those are the questions that define the mature AI economy.

## FIGURE

**The mature AI power stack — layers 1–3**

Physical and trust infrastructure: the substrate on which all subsequent AI value depends.

**1 Compute substrate**

AI still needs chips, memory, networking, power efficiency, and physical capacity even when the ownership story changes.

**STRENGTHENED**

- Chip suppliers
- Memory and packaging leaders
- Efficient infrastructure owners

**PRESSURED**

- Overbuilt capacity with weak utilisation
- Capital structures that assumed perfect demand timing

**2 Devices and operating systems**

More intelligence moves closer to the user, the permission boundary, and the default surface where work already happens.

**STRENGTHENED**

- Operating system owners
- Device ecosystems
- Default productivity surfaces

**PRESSURED**

- Pure model destinations with weak user surface
- Products that rely on users choosing them explicitly every time

**3 Identity and governance**

Action without permissions, policy, audit, and approval is operationally dangerous, so governance becomes a real product layer.

**STRENGTHENED**

- Identity providers
- Governance platforms
- Trusted control planes

**PRESSURED**

- Ungoverned agent tools
- Products that answer but cannot safely act

## FIGURE (CONTINUED)

**The mature AI power stack — layers 4–6**

Operating layers: where intelligence connects to data, authority, and economic action.

**4 Systems of record and context**

AI is only durable inside business systems when it can see authoritative data and act against trusted records.

**STRENGTHENED**

- Core systems of record
- Deeply integrated data platforms
- Trusted enterprise data owners

**PRESSURED**

- Generic workflow layers without record ownership
- Thin surfaces sitting above someone else's data

**5 Orchestration and execution**

Real AI work needs state, retries, human approval, long-running workflows, and auditability between intent and action.

**STRENGTHENED**

- Workflow orchestration layers
- Execution control planes
- Enterprise automation backbones

**PRESSURED**

- Prompt-only agent experiences
- Ad hoc automation with no durable state or governance

**6 AI-operated companies**

The deepest repricing may happen when companies use AI to change their own cost structure rather than merely buy AI features.

**STRENGTHENED**

- Small AI-operated businesses
- Companies with strong process knowledge
- Firms that can encode their operating model

**PRESSURED**

- Human-heavy service models
- Generic workflow SaaS
- Companies with no proprietary workflow or trust advantage

Layer in the mature AI economy	Why it matters	What gets rewarded after the crash
Compute, memory, power, networking	AI still needs physical substrate	Efficient, well-timed capacity and supply-chain leverage

Layer in the mature AI economy	Why it matters	What gets rewarded after the crash
Devices and operating systems	Intelligence moves closer to context and defaults	Ownership of user surface, permissions and routing
Models	Capability still matters, especially at the frontier	Specialized excellence or strategic platform position
Identity, governance and policy	Action without control is operationally dangerous	Trusted approval, audit, policy and control planes
Systems of record and data	AI needs authoritative context to act safely	Record ownership, data trust and embedded workflow leverage
Orchestration and execution	Real AI work needs retries, state, routing and human approval	Durable execution layers between intent and action
AI-operated companies	AI changes the cost structure of the firm itself	Lower coordination cost, faster iteration, thinner operating model

## Likely winner: chips, memory, packaging and power-efficient compute

The obvious winner group is still compute.

AI increases demand for specialized processing, memory bandwidth, packaging, networking and power efficiency. Even if centralized API economics weaken, compute does not stop mattering.

In fact, a more distributed AI world may increase the number of places where compute matters:

- frontier training;
- cloud inference;
- enterprise inference;

- local inference;
- edge devices;
- phones;
- laptops;
- cars;
- robots;
- industrial systems;
- data centre networking;
- memory and storage.

The specific winners may shift. Training-heavy demand is not the same as inference-heavy demand. Cloud GPUs are not the same as device NPUs. Memory bottlenecks are not the same as model architecture. Power efficiency becomes more important as AI becomes ambient, but the compute substrate remains strategic.

The caution is valuation and timing.

A company can be structurally important and still overvalued at the wrong point in the cycle.

That was true in dotcom. It will be true in AI.

## **Likely winner: devices and operating systems**

If intelligence moves closer to the user, devices and operating systems become more powerful.

This is why Apple, Microsoft, Google, Android, Samsung and other platform owners matter.

- The device has the context.
- The operating system has the permissions.
- The browser has the activity.
- The phone has the personal life.
- The laptop has the work.

The calendar, inbox, files, photos, messages, contacts, identity, location and app permissions all live close to the user.

A useful mainstream signal here is [Apple Intelligence](#), which is explicitly designed around on-device processing plus Private Cloud Compute. Even if Apple's particular implementation changes, the architectural direction matters: more intelligence moves closer to the user, the operating system and the permission boundary.

- A remote model API has intelligence.
- The device has life.
- That may be the more durable position.
- The mature AI user may not open a chatbot to ask for help. The help may already be inside the thing they are doing.
- That gives platform owners a strong position.
- They can route models. They can bundle features. They can privilege defaults. They can integrate privacy controls. They can decide which AI capabilities become part of the environment.
- This does not make them unbeatable.

But it gives them a structural advantage over pure model providers.

## Likely winner: systems of record

AI needs trusted context and that makes systems of record more important, not less.

Finance records. Customer records. Contracts. Orders. HR records. Identity. Product data. Risk data. Compliance records. Audit trails. Inventory. Usage history. Support history. Pricing rules. Legal obligations.

If AI is going to act, it needs to act on trusted data. The companies that own or govern those records have leverage.

This is why not all SaaS is equally exposed.

- A system of record can become more valuable if it becomes the trusted source for AI-operated workflows.
- A generic workflow layer around a system of record is more vulnerable and the distinction matters.

The mature AI economy may hollow out parts of SaaS while strengthening certain core platforms.

The record survives and the workflow around the record becomes fluid.

## **Likely winner: identity, permissions and governance**

AI without governance is a liability.

As AI systems move from answering questions to taking actions, permissions become central.

- Who is the AI acting for?
- What data can it access?
- What actions can it perform?
- Which systems can it write to?
- Which decisions require human approval?
- How is the action audited?
- How is responsibility assigned?
- How are failures detected?
- How are generated workflows tested?
- How is sensitive data prevented from leaking into the wrong model?

Microsoft is already productizing this layer directly through [Copilot Control System](#), which it describes as a way to secure, govern, manage and measure AI at enterprise scale. That does not prove Microsoft owns the mature AI economy. It does reinforce the broader claim that governance becomes a real product layer, not just an afterthought.

This makes identity and governance strategic.

The mature AI enterprise will need strong control planes. Not just model access. Not just chat interfaces. Not just prompt libraries.

Control planes.

The companies that provide identity, policy, audit, permissions, workflow governance, model routing, evaluation and observability may capture durable value.

This is one reason I think the model API moat is weaker than people assume.

The model can answer.

The governance layer decides whether the answer can become action and action is where business value lives.

## **Likely winner: workflow orchestration and durable execution**

AI agents are unreliable if they are just loops of prompts and tools.

Real business processes need durability, retries, auditability, state, compensation, human approval, error handling, observability and long-running execution.

This makes workflow orchestration important.

An AI-operated company or AI-native enterprise does not simply ask a model to do everything. It needs processes that can survive failure, wait for external events, retry safely, call tools, involve humans, record decisions and produce auditable outcomes.

That is not a chatbot problem. It is an execution problem.

The mature AI economy will need orchestration layers that sit between intent and action.

Those layers may become some of the most important infrastructure in the enterprise.

They will decide how AI work is broken into steps, which tools are called, what state is stored, what humans approve, what happens on failure and how outcomes are measured.

This is where AI becomes operations.

## **Likely winner: companies with proprietary process knowledge**

The best data is not always the largest dataset. Sometimes it is the most specific.

A company that deeply understands its own customers, pricing, risks, workflows, exceptions, supply chain, sales process, product constraints and regulatory environment has an advantage if it can encode that knowledge into AI-operated systems.

This is why domain expertise becomes more valuable, not less.

AI lowers the cost of producing software and output, but it does not automatically know what should exist.

The scarce knowledge becomes:

- what matters;
- what good looks like;
- what can go wrong;
- what exceptions exist;
- what rules are real;
- what customers actually need;
- what risk is acceptable;
- what trade-offs are tolerable.

Companies that can turn proprietary process knowledge into governed AI systems will be stronger than companies that merely buy generic AI features.

This is a hopeful part of the story.

AI does not only advantage the largest platform companies. It can also advantage companies that know their business deeply and can express that knowledge in software, data and workflows.

## **Pressured group: pure model API companies without distribution**

The model labs are important, but pure model API economics are vulnerable.

- If models become cheaper, smaller, local, open and specialized, then centralized API access loses some scarcity value.
- If devices and operating systems absorb everyday AI tasks, model labs lose user surface.
- If enterprises build private routing and governance layers, model labs become suppliers.
- If frontier capability remains expensive, only the hardest tasks justify premium pricing.

The strongest model companies will try to become platforms and they will build consumer products, enterprise tooling, developer ecosystems, agent frameworks, marketplaces, infrastructure partnerships and maybe devices.

Some will succeed, but the weaker ones, especially those without distribution, will struggle. The market currently treats model quality as if it can become durable ownership.

That is the questionable assumption.

## **Pressured group: thin AI wrappers**

Thin AI wrappers are the easiest casualties. A thin wrapper takes a model API, adds a narrow workflow and calls it a company.

Some wrappers are useful. Some will become real businesses by owning data, workflow, distribution or domain trust, but many will be features, not companies.

They will be copied by model providers, bundled into SaaS platforms, replaced by open source, generated internally by customers or undercut by broader tools.

This is normal in a platform shift.

The first wave produces many plausible ideas that do not have durable company-level economics.

When funding tightens, wrappers get exposed.

## **Pressured group: generic workflow SaaS**

SaaS is not dead.

But generic workflow SaaS is exposed.

- If the product is mainly forms, dashboards, routing, approvals, notifications, lightweight records and integrations, AI-generated internal software becomes a threat.
- If the pricing model depends on seat expansion, AI-driven labour compression becomes a threat.
- If the AI features are not differentiated, they become table stakes.
- If the product does not own the system of record, the data, the compliance model, the ecosystem or the customer relationship, it is easier to replace.

The weak SaaS company of the AI era is one that sells a generic workflow and calls it a platform.

The strong SaaS company owns a record, a network, a regulated function, a deep integration layer, a trusted operating surface or a governance position.

The distinction will become obvious after the bubble.

## **Pressured group: human-heavy services businesses**

The AI-operated company is a direct threat to human-heavy service models.

Agencies, consultancies, research shops, outsourcing firms, implementation partners, content operations, support services, sales operations, analytics services and many professional service niches will feel pressure.

The pressure will not be uniform. High-trust, high-judgement, high-relationship work survives longer.

Routine production, analysis, reporting, coordination and execution compress faster.

The dangerous competitor is not always a large incumbent with AI tools. It may be a small AI-operated company with lower costs, faster iteration and enough human expertise at the top.

That changes market structure.

## **Pressured group: companies with no proprietary workflow or data advantage**

AI rewards specificity.

A company with no proprietary data, no deep workflow, no customer trust, no distribution, no system of record, no governance role and no operational insight has little to defend.

This applies to startups and incumbents.

- Putting AI in the pitch does not create a moat.
- Adding AI features does not create a moat.
- Using a frontier model does not create a moat.

The question is what remains when everyone has access to similar intelligence.

If the answer is nothing, the company is exposed.

## The new strategic question for businesses

The most important question for businesses is not:

*How do we use AI?*

That question is too small.

The better question is:

*Which parts of our business become more valuable when intelligence is cheap, and which parts were only valuable because intelligence, software or coordination were expensive?*

That question cuts deeper.

It forces a company to examine its real moat.

- Do we own data?
- Do we own a workflow?
- Do we own trust?
- Do we own distribution?
- Do we own a regulated position?
- Do we own a system of record?
- Do we know something specific about our customers that others do not?
- Can we encode our process knowledge into software?
- Can we operate with fewer coordination layers?
- Can we build custom internal systems faster than before?
- Can we expose our capabilities as APIs and tools?
- Can we govern AI safely enough to let it act?

These are the questions that matter after the bubble.

## The investor question

The investor version is similar. Do not ask only whether a company has AI exposure.

Ask whether it captures AI value.

- Is AI reducing cost or increasing spend?
- Is AI increasing margin or subsidizing usage?
- Is the company selling a durable capability or a temporary wrapper?
- Is the company protected by data, distribution, workflow, governance, compute or trust?
- Is the company exposed to model commoditization?
- Is the company exposed to seat compression?
- Is the company's capex ahead of demand?
- Is the company the owner of the workflow or a supplier to someone else's workflow?

During the bubble, exposure is enough. After the bubble, capture matters.

## The final prediction

The AI bubble will burst because AI is real.

It will burst because AI is real enough to make models smaller, cheaper and more local. It will burst because AI is real enough to make custom software cheaper. It will burst because AI is real enough to pressure generic SaaS. It will burst because AI is real enough to create AI-operated companies with radically different cost structures. It will burst because AI is real enough to force investors to ask who actually owns the future.

The market has correctly sensed a paradigm shift but it has not yet correctly mapped the ownership layer.

That is the mistake.

When belief changes, the bubble breaks. After that, the real AI economy begins.

It will be useful. It will be profitable. It will be more centralized than many hope. It will be more distributed than some incumbents want. It will be more extractive than the early optimism suggests. It will compress labour, software and coordination. It will create new company forms. It will destroy some moats and create others.

It will not look exactly like the first story. The first story almost never survives contact with the mature paradigm.

I still have those business cards from the dotcom bubble. They remind me that people can stand close to the future and still be wrong about who owns it.

That is where I think we are with AI.

The future is real.

The current map is wrong.

If you want the argument from the top again, return to the [preface](#) and read forward through the series as one continuous thesis.