

THE NEW ECONOMICS OF INTELLIGENCE

All Roads Lead to *Causum*

Why the economics of AI ends in the economics of justified machine cognition

COMPUTE

TOKENS

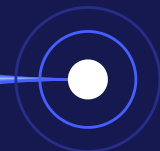
AGENTS

DECISIONS

HARNESS

GOVERNANCE

ASSURANCE



Intelligence is getting cheaper. Justified intelligence is getting more valuable.

CAUSUM
POSITION PAPER

An argument in four parts

Abstract

I

The Question Has Changed

- 01 The Paradox of Cheap Intelligence
- 02 From Token Economics to Intelligence Economics
- 03 The Decision Dividend — and Its Missing Variable

II

Delegation and Justification

- 04 Delegated Cognition
- 05 Justification: The Missing Economic Variable
- 06 Four Thresholds of Machine Work
- 07 The Economics of Authority

III

The Economics of Justification

- 08 Where the Harness Ends
- 09 From Token Optimization to Justification Optimization
- 10 The Delegation Frontier
- 11 More Than AI Governance
- 12 The New Economic Stack

IV

Convergence

- 13 All Roads Lead to Causum
- 14 How Causum Produces Justification
- 15 The Causum Thesis

Conclusion · Evidence Position · Sources

ABSTRACT

The economics of artificial intelligence is changing its unit of account. For most of the generative era, AI was priced and argued in terms of compute, inference, tokens, labour substitution, and productivity. In 2026, mainstream executive analysis — most visibly McKinsey's *New Economics of AI* series — began to say something different: unit prices are collapsing, total bills are rising, and AI demand should be managed around *outcomes*, not tokens.

That is a real maturation. This paper argues it stops one step short. Between machine output and economic outcome sits a decision that no cost model yet prices: *may this institution rely on what the machine concluded?* We call the answer to that question **justification**, and we argue it is the variable that ultimately determines how much intelligence is worth.

THE PROPOSITION

Causum is the economics of *justified machine cognition*.

Causum operates in the layer between machine cognition and institutional action — the point at which machine-generated reasoning is grounded, evidenced, auditable, and authorized enough that an organization can rationally delegate consequential work to it.

The title is a claim about convergence, not inevitability. Compute economics, token economics, agentic economics, decision economics, harness engineering, governance, and assurance are separate roads. Each, followed far enough, arrives at the same unresolved question: *when is machine cognition justified enough to be delegated?*

THE OLD QUESTION

How much does intelligence cost?



THE CAUSUM QUESTION

How much is *justified* intelligence worth?

The Question Has Changed

01

The Paradox of Cheap Intelligence

The first economic question of AI was simple: *how much does intelligence cost?* The industry answered in GPUs, inference, and price per million tokens — and by that measure, intelligence has become dramatically cheaper. In 2023, GPT-4's 8K-context API was priced at \$60 per million output tokens; models of comparable benchmark capability now run for a fraction of that. [Fortune, 2026]

Yet enterprise AI bills keep rising. As the unit cost of intelligence collapses, consumption explodes — above all in agentic systems that reason, retry, and refine across many steps.

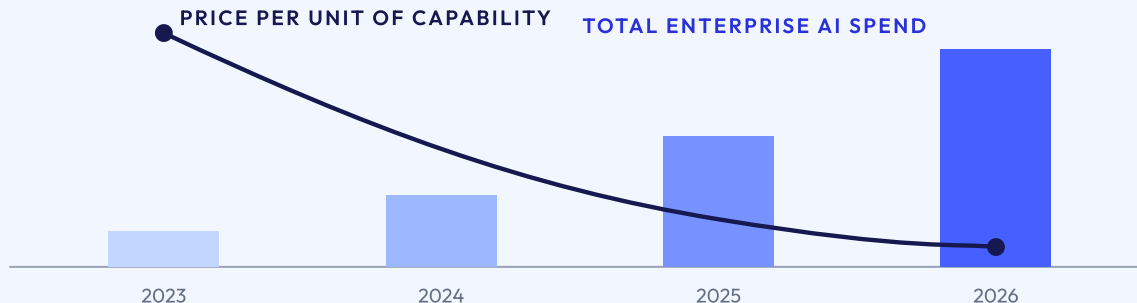


Fig. 1 — The paradox of cheap intelligence. Illustrative shape, not to scale.

93%

of surveyed enterprises report exceeding their AI budgets

~4×

rise in AI spend moving from isolated use cases to enterprise-wide adoption

30×

possible cost variance for the same agentic task between runs

1 in 5

organizations say AI spend is beginning to constrain operating costs

Fig. 2 — The bill arrives. Sources: McKinsey Enterprise AI FinOps survey, May 2026 (n=75 qualified); McKinsey State of AI 2026, as reported by Fortune and Business Insider, Sept 2026.

So the question has changed. Not *how cheap can intelligence become?* but **what is it worth to let a machine perform cognition on an organization's behalf?**

McKinsey's response is to manage AI demand around business outcomes rather than token consumption. That moves the unit of analysis from the *token* to the *task*, and from the task to the *value of the resulting action*. It also exposes a problem the outcome lens alone cannot solve: a machine can reach a decision quickly and cheaply without reaching a decision an organization should trust.

That gap is where Causum begins.

02

From Token Economics to Intelligence Economics

A token price is an input metric. Business value is an output metric. Between them runs an increasingly long production process — and the token is only its first link.

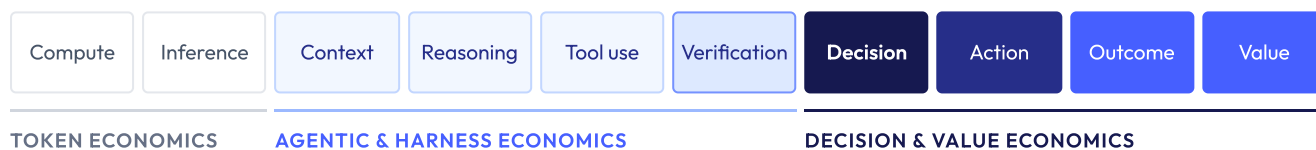


Fig. 3 — The production chain of intelligence. Each economic discipline has claimed one stretch of it.

McKinsey's 2026 *New Economics of AI* collection sets the cost of intelligence, agentic workflows, the decision dividend, inference economics, and AI value creation side by side. That convergence is itself the signal: AI is acquiring its own economic theory, and its scarce resource is no longer compute. It is **useful intelligence** — which is not the same thing as raw model output.

03

The Decision Dividend — and Its Missing Variable

McKinsey's analysis of the *decision dividend* makes the bridge explicit: AI's largest gains may come less from replacing labour than from helping organizations make better and faster decisions — raising decision throughput, asset utilization, and the capture of opportunity.

OLD EQUATION

~~AI value = labour eliminated~~

DECISION DIVIDEND

quality × speed × volume × execution × opportunity
captured

WITH CAUSUM

quality × speed × volume × consequence × **justification**

Fig. 4 — Three equations for AI value. Justification is the term that converts a fast decision into a delegable one.

The missing variable is **justification**. A decision dividend can only be collected on decisions an organization is willing to act on. And the more consequential the decision, the more its economic value depends on establishing whether the machine's reasoning deserves reliance. Multiply a brilliant decision by zero justification and the dividend is zero.

II

PART TWO

Delegation and Justification

04

Delegated Cognition

Agentic AI changes the relationship between people and machines. Conventional software executes a specified instruction. A generative model answers a request. An agent does something new: *it determines what needs to be done, then performs a sequence of cognitive and operational steps on another party's behalf.*

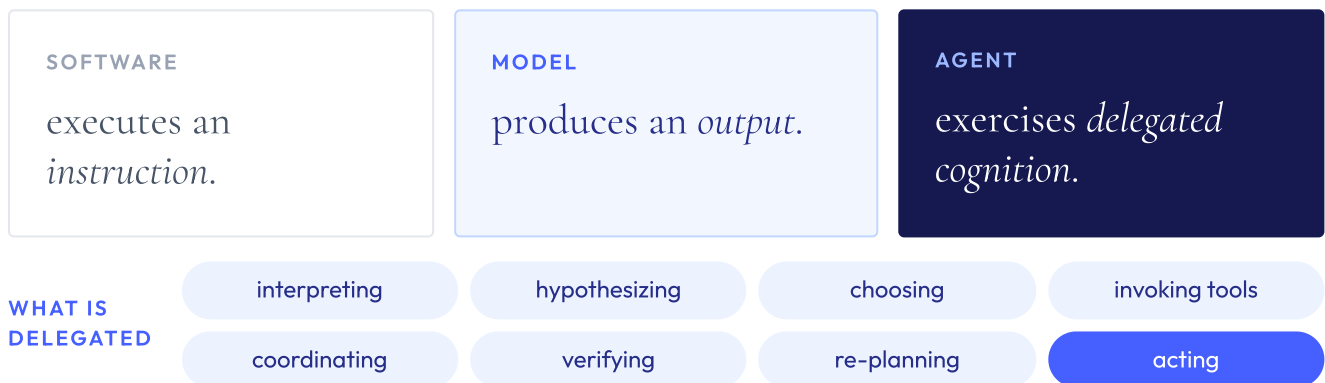


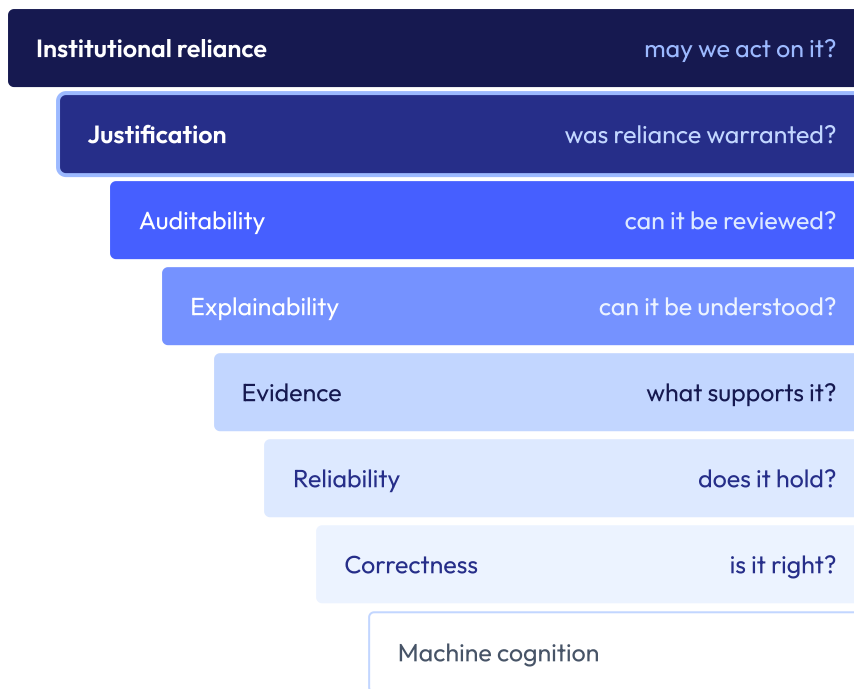
Fig. 5 — Three relationships between an organization and a machine.

The organization is no longer buying software. It is buying access to *cognition* — and that changes the cost structure of organizational decision-making. The economic question becomes: **how much cognition should an organization delegate to a machine, under what conditions, and with what authority?**

05

Justification: The Missing Economic Variable

A machine can produce an answer. It can produce a good answer, even a highly probable one. None of that establishes that an institution should *rely* on it. Between machine cognition and institutional reliance lies a ladder — and each rung adds economic value.



Most AI evaluation stops at the bottom three rungs. The economic value is unlocked at the top two.

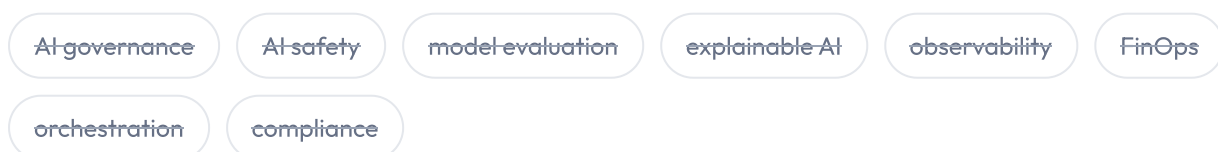
Fig. 6 — The justification ladder.

The last rung is the economically decisive one. An output that cannot be justified has limited utility even when it is usually right, because someone must still check it, absorb its risk, or refuse to act on it. A system that produces *justified* cognition lets an organization delegate progressively more consequential work.

The economic value of machine cognition is set not by its accuracy or its cost, but by how much consequential work can rationally be delegated to it.

That is the territory Causum occupies. Stated formally: **Causum establishes the conditions under which machine-generated cognition acquires enough justification to support delegated organizational action.**

CAUSUM IS NOT SIMPLY



Each solves part of the problem. Causum addresses their *economic consequence*: what becomes possible to delegate once machine cognition can be justified.

Four Thresholds of Machine Work

McKinsey's agentic-economics work observes that the same task can have radically different economics depending on how machine work is executed — and that organizations that govern machine work well extract more value from the same intelligence spend. That suggests four distinct thresholds, each an economic step change.



Fig. 7 — Each threshold unlocks more delegable value. Only the last survives an audit, a dispute, or a regulator.

The final threshold matters more as the consequences of delegation rise. Reliable work is enough for a draft. Justified work is what a board, a counterparty, an insurer, or a regulator will ask for when the decision goes wrong.

The Economics of Authority

Organizations have always delegated authority — because human institutions built mechanisms that made delegation justifiable. AI introduces something without precedent: *cognitive labour can now be delegated without transferring authority to another accountable human.*

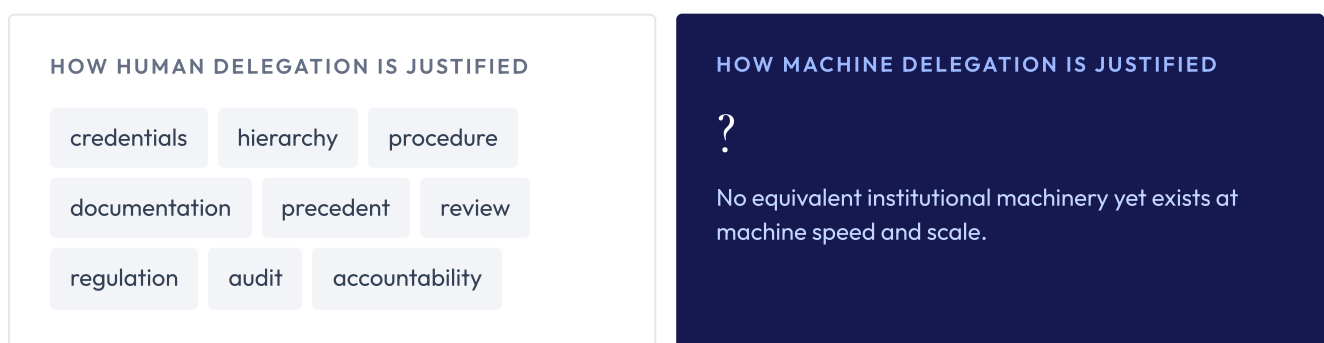


Fig. 8 — Centuries of machinery for justifying human delegation; almost none for machine delegation.

So who — or what — justifies the machine's decision? It cannot be “*the model was accurate,*” “*the model was highly capable,*” or “*the model scored well on a benchmark.*” Each describes the machine in general. None addresses the decision in particular. The economically relevant question is far more demanding:

Was *this* machine cognition sufficiently justified for *this* decision, under *these* conditions, with *these* consequences?

III The Economics of Justification

08

Where the Harness Ends

The model is not the whole cognitive system. The **AI harness** — memory, context, planning, tool selection, execution, verification, state, recovery, permissions, integration — turns model capability into operational machine work. This is why model benchmarks alone no longer measure agentic capability: the same model in two harnesses is two different workers. It is also why the same task can vary in cost by an order of magnitude.

But the harness also reveals the next layer. It decides *how* machine cognition is operationalized. It cannot decide *whether* that cognition deserves reliance — because the producer of the work cannot be the final judge of it.

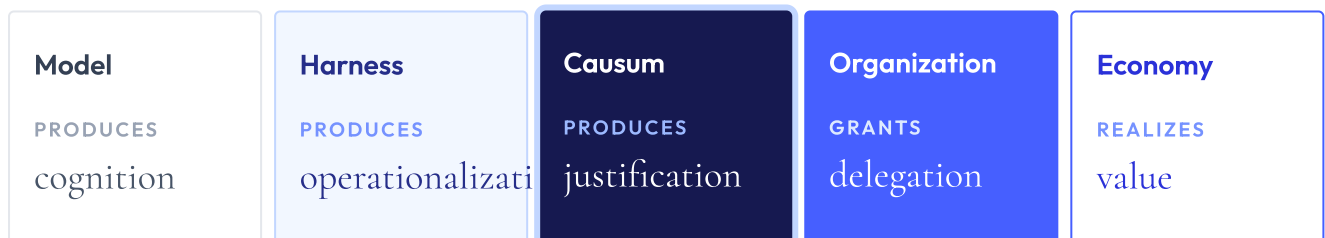
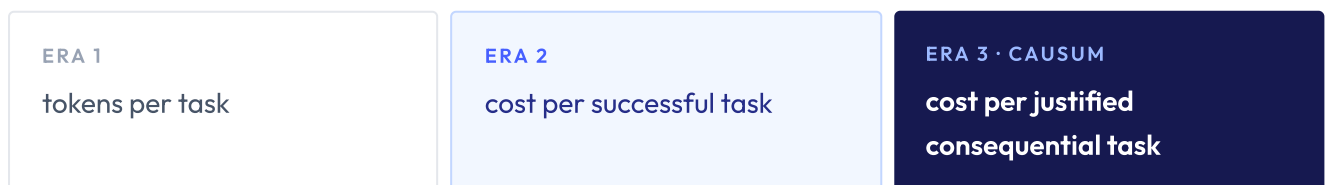


Fig. 9 — Five layers, five products. Delegation is only granted on what has been justified.

09

From Token Optimization to Justification Optimization

The industry first optimized *tokens per task*. It is now optimizing *cost per successful task*. The next metric is the one that decides how far delegation can go:



A worked example shows why the third metric reverses conclusions drawn from the first. Two agents complete the same consequential task.

	AGENT A	AGENT B
Cost of cognition	\$0.10	\$0.40
Success rate	90%	95%
Evidence produced	minimal	auditable trail
Human verification	every task · \$4.00	10% sample · \$0.40
Token economics favours	A — 4× cheaper	—
Cost per justified task	\$4.56	\$0.84

Fig. 10 — Illustrative. Cost per justified task = (cognition + verification) ÷ success rate. Verification assumes ~5 minutes of expert review at \$48/hour.

Agent A is four times cheaper to run and more than five times more expensive to rely on. The difference is not intelligence. It is justification — and the human cost of compensating for its absence. The full economics of delegation can be written as:

THE VALUE OF DELEGATION

value of autonomous action

- cost of cognition
- cost of verification
- cost of failure
- cost of residual uncertainty
- cost of insufficient justification

THE CAUSUM TERM

IT INCLUDES

escalation

duplicated review

regulatory exposure

operational risk

contractual liability

reputational damage

the inability to delegate at all

Fig. 11 — The delegation equation. Every other term is already measured somewhere; the last one is not.

The Delegation Frontier

Every organization has an implicit **delegation frontier**. Below it, machine cognition can be delegated economically. Above it, human intervention remains necessary. The frontier is not set by model capability alone; it is the product of capability, evidence, reliability, context, governance, and consequence.

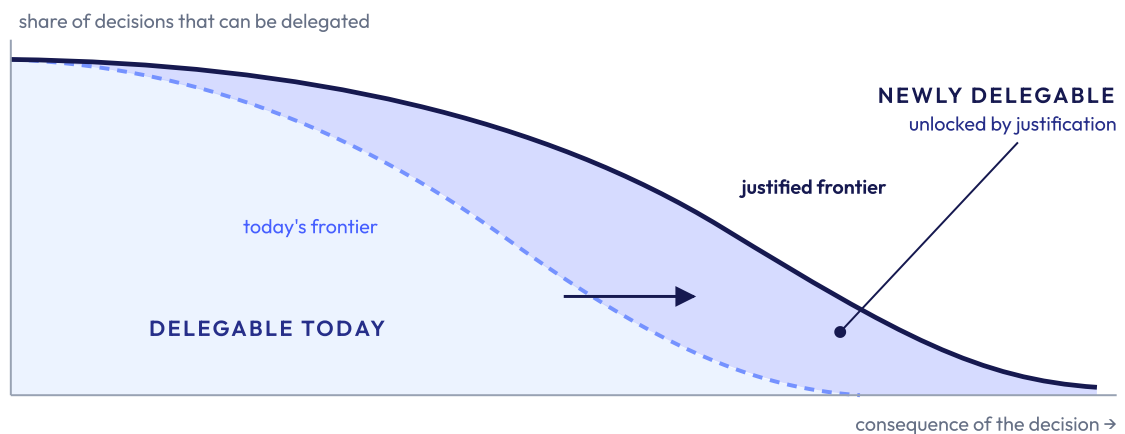


Fig. 12 — The delegation frontier. Justification moves it outward, precisely where decisions are most consequential — and most valuable.

As systems get better at justifying their outputs, the frontier moves outward. The gain is concentrated on the right of the curve — in high-consequence decisions where the economic stakes are largest and today's delegation is weakest.

Better justification increases the amount of economically delegable cognition.

More Than AI Governance

Each adjacent discipline asks a necessary question. Causum asks what its answer is *worth*.

DISCIPLINE	ASKS	CAUSUM ASKS
Governance	Is the system compliant?	How much cognition can we rationally delegate?
Evaluation	How well does the model perform?	What consequential work does that performance justify?
Observability	What happened?	Does what happened establish grounds for reliance?
FinOps	How much did the AI cost?	What justified cognition did that spend buy — and what authority did it unlock?

Fig. 13 — Four disciplines, four questions, one economic consequence.

12

The New Economic Stack

Assemble the pieces and the AI economy resolves into a stack. The industry has spent a decade and hundreds of billions of dollars optimizing its lower layers. The economic opportunity is moving up.

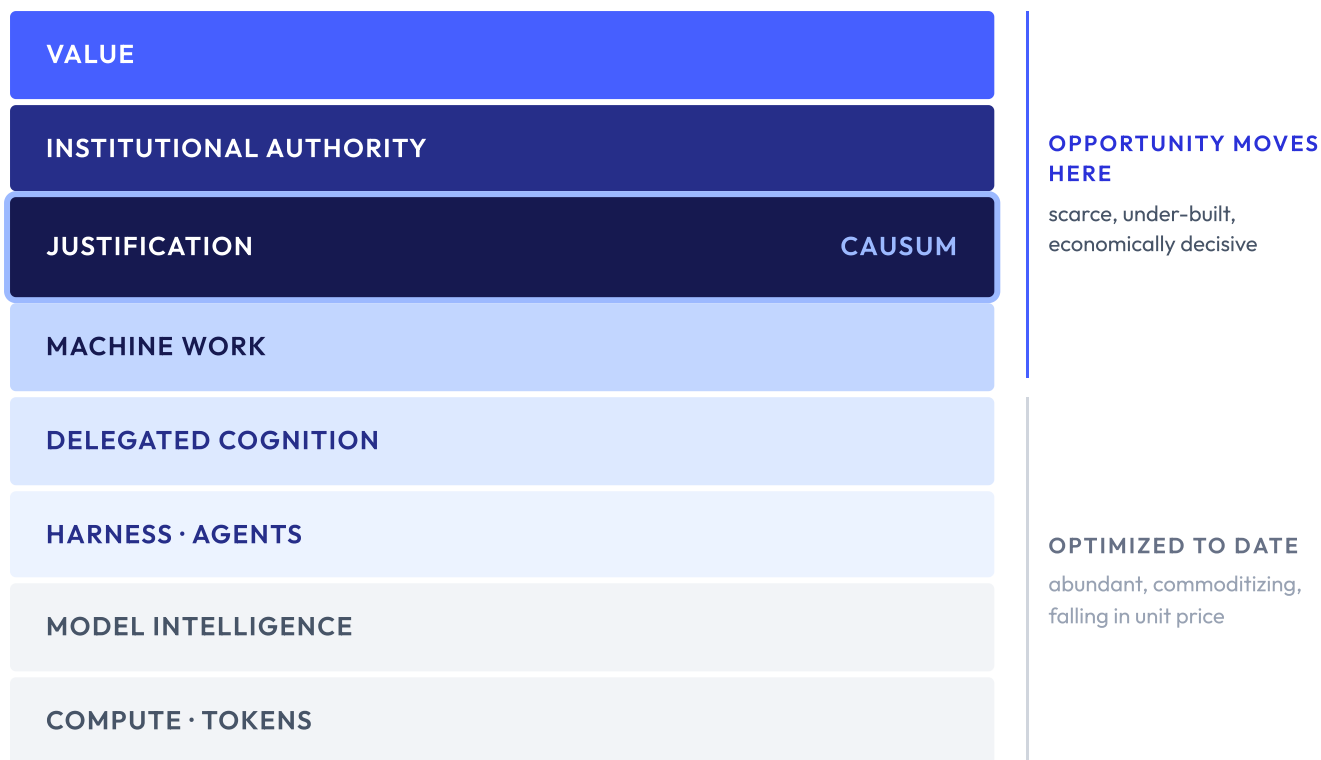


Fig. 14 — The new economic stack of AI.

McKinsey's 2026 work is significant because it begins this ascent — from token consumption to intelligence consumption, from labour savings to decisions, from deployment to outcomes. *Causum extends the trajectory one layer further: to the layer that determines how much authority the layers beneath it can carry.*

IV

PART FOUR

Convergence

13

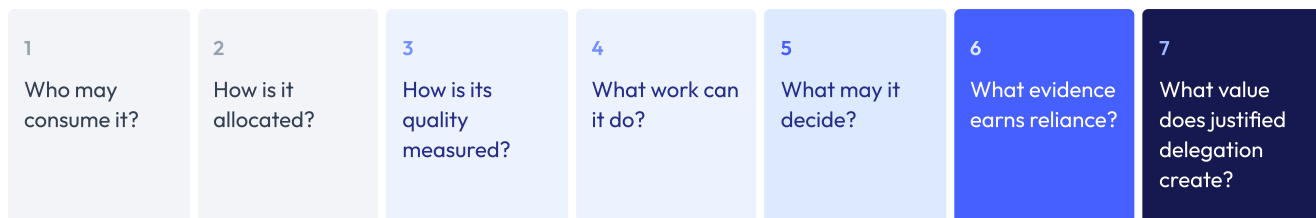
All Roads Lead to Causum

The title is deliberately provocative. Not every theory of AI economics is secretly about Causum. But seven independent developments — each pursued by different people, for different reasons — converge on the same unresolved problem.

COMPUTE	Cheaper inference means more intelligence consumed.	WHERE THEY MEET When is machine cognition justified enough to become economically delegable? THE CAUSUM QUESTION
TOKENS	Cheaper tokens make token count a weaker proxy for value.	
AGENTS	More machine work demands an economics of autonomous execution.	
DECISIONS	Decision quality and throughput become economic variables.	
HARNESS	Orchestration and verification decide what models can do.	
GOVERNANCE	Consequential action requires accountability and control.	
ASSURANCE	Specific outputs must evidence reliance.	

Fig. 15 — Seven roads, one destination.

Mainstream executive discourse now treats **intelligence itself as an economic resource** — something organizations consume, allocate, and manage against outcomes. Once intelligence is a resource, seven questions follow in sequence:



INFRASTRUCTURE OF THE AI-NATIVE ORGANIZATION

ECONOMICS

The first six define the infrastructure of an AI-native organization. The seventh defines its economics — and it cannot be answered without the sixth.

14 How Causum Produces Justification

Justification is not a sentiment or a score. It is an artifact that can be constructed, inspected, and challenged. Causum builds it from four components and resolves it into one of three determinations.



Fig. 16 — From components to determination. Each outcome moves the delegation frontier.

This is where the economics becomes operational. CONFORM lets delegation scale without proportional review. DIVERGE prevents unjustified reliance before it becomes liability. REFER converts the most expensive input in the system — expert human judgment — from a blanket cost on every task into a targeted cost on the few that need it.

The Causum Thesis

Models will keep getting more capable. Inference will keep getting cheaper. Agents will run longer, more complex work. Harnesses will become more sophisticated. Each of these trends is real, and each pushes the economic bottleneck upward.

The limiting factor will increasingly be an organization's willingness — and ability — to trust machine cognition with consequential authority. That is not merely an engineering problem. It is an economic one.

The scarce resource in advanced AI will not be intelligence. It will be *justified intelligence* — intelligence that can be safely and economically delegated.

CONCLUSION

YESTERDAY

How cheap can we make AI?

TODAY

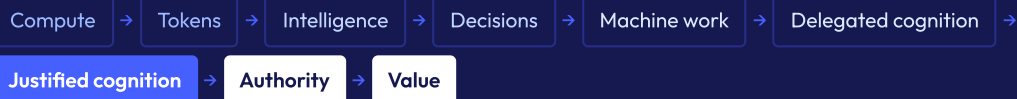
How much value do we get from the intelligence we consume?

NEXT

How much consequential work can we delegate to machines?

CAUSUM

How much of that delegation can be justified?



The first generation of AI economics optimized the start of this chain. The agentic economy is moving through its middle. **Causum** occupies the transition from cognition to justified authority.

If intelligence is the resource, agents are the workers, harnesses are the machinery, and decisions are the units of organizational action, then justification is the mechanism that determines how much authority can economically be delegated to machines.

Causum is the economics of *justified machine cognition.*

That is the next economic frontier of AI — and where Causum begins.

What is reported, and what is argued

Reported

Falling unit prices alongside rising enterprise AI spend; budget overruns; high cost variance in agentic tasks; McKinsey's recommendation to manage AI demand around outcomes.

Illustrative

The Agent A / Agent B worked example, the shape of Fig. 1, and the delegation frontier in Fig. 12 are models, not measurements.

Causum thesis

Justification is the missing economic variable of AI, and Causum is the economics of justified machine cognition. This paper builds on McKinsey's analysis; it is not endorsed by it.

SOURCES

- 1 McKinsey & Company — *The New Economics of AI*, 2026 collection, incl. *The Rising Cost of Intelligence* (Oct 2026)
- 2 McKinsey — *The cost of intelligence: How CIOs can manage AI demand at scale* (Jul 2026)
- 3 McKinsey — *Is that AI agent worth it? Agentic economics and the modern operating model* (Jul 2026)
- 4 McKinsey — *The Decision Dividend: How AI Creates Economic Value* (Aug 2026)
- 5 Fortune — *McKinsey: Cheaper AI models, bigger AI bills* (Sept 2026)
- 6 Business Insider — *McKinsey warns that AI may get a lot more expensive for companies* (Sept 2026)
- 7 Financial Times — *Why Big Tech Wants More Economists* (Oct 2026)