1337

ADVISORS

The Economics of Timing

Intelligence Manufacturing, Capital Formation, and the Evolution of Compute Markets

Mike Ossowski
1337 Advisors

Contents

- **Part I · The Manufacturing Lens**
  - 1. Reconsidering What Compute Represents
  - 2. The Central Mismatch: Production Capability and Economic Realization
  - 3. The Different Economic Lives of Machinery and Factories
- **Part II · The Financial Consequences of Different Clocks**
  - 4. What Stanley Lee's Credit Analysis Reveals
  - 5. Hyperbolic and the Economics of Term Transformation
  - 6. A Timing Example: When the Same Demand Produces Different Credit Outcomes
- **Part III · The Market That Must Still Be Built**
  - 7. Price Discovery Is Not Value Discovery
  - 8. Why Intelligence Demand Need Not Become GPU Scarcity
  - 9. Where Scarcity May Migrate
  - 10. The Measurement Infrastructure: From Benchmarks to Bankable Economics
- Conclusion

© October 2026 1337 Advisors. All rights reserved.

Executive Summary

The ability to manufacture intelligence is advancing on a different timeline from the economy's ability to realize its value. Meanwhile, the capital required to build that manufacturing capacity must be committed years before either trajectory is fully understood. This separation of technological capability, economic realization, and financial obligation is emerging as a defining feature of the compute economy.

The consequences extend beyond whether AI demand ultimately proves large enough. They determine which infrastructure assets retain value, which capital structures survive, where scarcity emerges, and how financial markets develop to transfer the resulting risks.

This misalignment creates exposures that cannot be understood through aggregate AI demand forecasts, GPU rental prices, or hardware performance improvements considered separately. The issue is not that demand must disappoint, nor that investment must be excessive. It is that the timing of useful demand, technological substitution, cash receipts, and financial obligations can determine outcomes even when long-run expectations prove broadly correct.

A manufacturing lens clarifies the problem. A GPU-hour measures access to a machine, not the cost of producing a useful result and not the economic value of that result to its consumer. GPUs are productive machinery; power-secured data centers are part of the factory; models, data, software, and operations shape the production process. Their economic lives and sources of scarcity differ. A new accelerator can undermine the earning power of installed hardware while increasing the value of an adaptable facility, or efficiency gains can reduce the value of both. Neither result follows automatically.

On the demand side, the technological frontier and the economic realization frontier are distinct. An organization may be able to use a model before it can deploy it reliably, integrate it into a workflow, and earn an acceptable return. Firms with established computational operations, clean feedback loops, and measurable outcomes may convert new capabilities into value sooner than firms requiring organizational transformation. Adoption is therefore uneven, not a single economy-wide clock.

Two contemporary analyses illuminate the financial consequences without themselves advancing this paper's overarching thesis. Stanley Lee describes a compute credit market beginning to differentiate more carefully among offtakers, operators, collateral, and recovery pathways. Hyperbolic Labs proposes purchasing long-duration compute capacity and distributing it in

© October 2026 1337 Advisors. All rights reserved.

shorter-duration commitments, with a modeled price hedge against part of the tail risk. Both illustrate the need to understand when cash flows arrive and which risks survive the contract or hedge.

The implication is practical: a mature compute market requires more than reliable GPU-hour quotations. It requires comparable workload economics, transparent distinctions among hardware risk, operating-cluster risk and facility risk, credible recovery scenarios, and instruments whose specifications actually match the exposures being hedged. Price discovery is necessary. Value discovery is the harder and as yet unfinished task.

This is a conceptual economic framework, not a forecast that compute is overbuilt, that AI demand is insufficient, or that any particular asset price must fall.

Part I · The Manufacturing Lens

Compute is access to a specified resource. It is also manufacturing capacity for a specific kind of output. The distinction determines what gets priced, what gets hedged, and what gets missed.

---

1. Reconsidering What Compute Represents

Compute is commonly sold as access to a specified resource for a specified period. The GPU-hour is an effective commercial unit: it helps buyers compare offers, helps suppliers quote capacity, and can support standardized financial contracts. But the convention describes access to productive machinery, not the economics of what that machinery produces.

A more revealing frame is manufacturing. Accelerators, memory, interconnect, networking, software, cooling, and electricity make production possible. Models, data, algorithms, and operating practices determine how the equipment is used. Training creates or improves model capabilities; inference applies models to generate predictions, classifications, text, code, simulations, or actions. These outputs may contain useful intelligence, but usefulness is task-dependent. Even useful intelligence has no guaranteed monetary value until someone can employ it productively.

Four layers must therefore remain separate:

© October 2026 1337 Advisors. All rights reserved.

Layer	Economic question	Illustrative measure
Machine access	What does it cost to obtain computational resources?	$/GPU-hour, with defined configuration and service terms
Computational output	What does the system generate?	Tokens, predictions, completed jobs, simulations
Useful output	Does the result meet the task's required quality, speed, and reliability?	Quality-adjusted tasks completed per unit of capacity
Realized economic value	What incremental benefit can the customer actually capture?	Revenue contribution, avoided cost, risk reduction, or improved decisions

Exhibit 1: Four layers of compute economics and their corresponding questions.
Source: 1337 Advisors.

The distinction matters because a lower hourly price does not necessarily imply a lower production cost. A system that rents at $2 per hour and completes a comparable task in two hours costs $4 in machine access. Another that rents at $3 per hour but completes it in forty minutes costs $2. Those figures ignore other operating costs, but they show why the rental rate alone cannot establish competitiveness. Comparisons must hold constant the model, output quality, latency, batching, memory footprint, topology, software stack, and delivery reliability.

Nor does a lower cost per useful result establish the result's value. A prediction can be accurate but economically irrelevant; an automated action can be useful but costly to supervise; a model may create value only when combined with proprietary data, human judgment, or distribution. The value of manufactured intelligence is partly determined outside the factory.

> KEY INSIGHT
>
> The fundamental distinction is between the price of the machine-hour, the cost of the useful output, and the value realized from consuming that output. Markets that improve the first measure do not automatically solve the other two. Financial infrastructure built entirely around access pricing leaves the more economically consequential questions unanswered.

© October 2026 1337 Advisors. All rights reserved.

2. The Central Mismatch: Production Capability and Economic Realization

Building compute capacity requires commitments before its eventual economics are known. Hardware must be ordered, sites secured, power and interconnection arranged, financing obtained, and customers contracted. Meanwhile the usefulness of the resulting output evolves through research, model architecture, software optimization, and changes in how customers work.

This creates two related but distinct frontiers. The technological frontier describes what systems can accomplish under specified conditions. The economic realization frontier describes what organizations can deploy reliably and profitably after accounting for integration, data, governance, workflow change, supervision, and competitive alternatives. Progress at the first frontier can precede progress at the second. The reverse is also possible: a mature application can create demand for capabilities that existing hardware already provides economically.

The gap is not uniform. Quantitative trading firms offer an instructive hypothesis, not a proven claim about any named firm's returns. Organizations already built around data, prediction, optimization, simulation, and frequent performance measurement may be able to evaluate and monetize some AI improvements faster than enterprises still redesigning their operating processes. The relevant advantage is not simply access to powerful models. It is the capacity to test whether an incremental capability improves decisions and to act on that evidence quickly. Even there, competition may rapidly transfer the gains to customers or counterparties rather than preserve excess profits for producers.

The resulting demand relationship is deceptively simple:

Required machine capacity ≈ demanded useful output ÷ useful output per unit of machine capacity

This is an organizing identity, not a directly measurable universal formula. Useful output must be defined for each workload, with appropriate quality and service requirements. The numerator changes as applications become economically attractive. The denominator changes with accelerators, models, algorithms, scheduling, software, and system design. Improvements in efficiency can reduce machine-hours per task; lower costs can also stimulate new uses and expand total task volume. Which effect dominates is empirical and may differ across workloads and time periods.

© October 2026 1337 Advisors. All rights reserved.

A surge in AI use is therefore not a sufficient forecast of GPU-hour demand. Nor is a surge in token production necessarily a surge in useful intelligence or in realized economic value. The critical question is how quickly the economy expands economically viable uses relative to how quickly the manufacturing system increases its useful output per unit of capacity.

> KEY INSIGHT
>
> Timing is the variable the formula cannot suppress. Technological possibility can arrive before the customer's ability to capture its value. That gap becomes financially consequential when it is measured against fixed obligations, because contracts and debt service do not wait for adoption.

3. The Different Economic Lives of Machinery and Factories

A GPU, an operating cluster with customer contracts, and a power-secured data center are economically connected but distinct assets. They should not be assigned the same depreciation logic, collateral assumptions, or recovery pathways.

A GPU's competitive life depends on the cost at which it can continue to produce relevant workloads relative to feasible substitutes. Newer hardware may be faster yet offer little advantage for some tasks; older hardware may remain technically functional but become uneconomic when newer systems produce equivalent output at lower total cost. Technological obsolescence and economic obsolescence are not synonymous. Neither is physical service life equivalent to the period over which an asset can earn an attractive return.

The facility has a different exposure. Its value depends on power access, interconnection, cooling, networking, physical location, regulatory conditions, and adaptability to successive equipment generations. A site that can accommodate more demanding equipment may remain valuable through several GPU cycles. But power access is not a guaranteed scarcity rent: competing sites can emerge, workloads can migrate, and higher efficiency can reduce demand for some physical capacity.

An operating cluster introduces yet another layer: service delivery, software integration, customer contracts, and operator competence. A long-term take-or-pay agreement may support financing, but its value depends on the operator's ability to meet service commitments and on the enforceability and transferability of the contractual cash flows.

© October 2026 1337 Advisors. All rights reserved.

For underwriting, at least three valuation questions must be kept separate: value in continued operation, value after redeployment or re-letting, and liquidation or recovery proceeds under stress. They may diverge sharply when a borrower fails. A reference rental index can inform these estimates; it cannot determine them without assumptions about utilization, operating costs, workload substitution, equipment condition, location, contract rights, and time to redeploy.

> KEY INSIGHT
>
> A loan secured by accelerator hardware, a loan against an operating cluster with enforceable customer contracts, and financing for a power-secured facility are not interchangeable credit exposures. Their cash flows, substitution risks, recovery pathways, and useful lives differ in ways that matter enormously under stress. Credit structures that treat them as equivalent are pricing something other than the risk they have actually taken on.

Part II · The Financial Consequences of Different Clocks

Two recent market documents examine the timing problem from different positions. Neither frames it as this paper does. Both provide evidence that the framing is necessary.

4. What Stanley Lee's Credit Analysis Reveals

In "The compute credit cycle just turned, and why that's a good thing," Stanley Lee argues that a changing compute credit market is beginning to price distinctions it previously treated less explicitly. He identifies several drivers of wider financing spreads: changes in benchmark rates, issuer-specific risk, issuance concentration and investor capacity, and a more durable reassessment of operators, collateral, and recovery. His central claim is not that compute demand has disappeared. It is that the offtaker contract remains central to the credit but is no longer the entire underwriting story. [1]

Lee explicitly rejects a simple comparison with fiber overbuilding, emphasizing that capacity can be contracted at delivery even as financing conditions tighten. This paper does not attribute a demand-timing thesis to him. His analysis is useful here because it shows where timing becomes

© October 2026 1337 Advisors. All rights reserved.

visible in credit: the operator must finance and deliver the asset, the customer must remain creditworthy and willing to perform, and the collateral must retain recoverable value at the moment a lender might need it.

A contract with a strong customer is not equivalent to an unconditional cash flow. Financing can fail before deployment. Service-level failures can reduce collections or trigger remedies. A concentrated set of offtakers can exhaust lenders' capacity to add exposure even when each individual contract looks attractive. Collateral that is valuable as part of a functioning cluster may be substantially less valuable when it must be removed, reconfigured, remarketed, or re-let under stress.

Secondary brokerage and bilateral transfers of physical compute capacity are emerging. What remains limited is the depth, standardization, transparency, and reliable transferability needed to diversify and hedge exposures at institutional scale. A lender's inability to reduce a concentrated position is not evidence that every underlying project lacks demand. It is evidence that the market for transferring the risk is less developed than the market for creating it.

Lee's distinction is consistent with the manufacturing lens: lenders increasingly need to assess not just who promised to buy the output, but whether the machinery and factory can deliver it and what remains recoverable if they cannot. That is an interpretation developed here, not a claim Lee makes about the timing of economy-wide AI adoption. The particular market yields, financing spreads, and anecdotes in his article are his observations and have not been independently verified in this paper. [1]

---

5. Hyperbolic and the Economics of Term Transformation

Hyperbolic Labs' September 2026 Compute Market-Making White Paper addresses another mismatch: infrastructure owners and lenders often prefer long-duration contracted revenue, while some customers prefer shorter, more flexible commitments. An intermediary that buys long-dated capacity and sells shorter-duration access can earn a premium for absorbing that mismatch. [2]

Hyperbolic illustrates the idea with a modeled 36-month commitment for 256 B300 GPUs at $4.30 per GPU-hour. Assuming 24-hour availability over 365 days per year, the nominal commitment is approximately $28.93 million. A 20% deposit is approximately $5.79 million. The deposit is not the full economic exposure: the intermediary remains responsible for the remaining

© October 2026 1337 Advisors. All rights reserved.

contractual payments and the liquidity required to carry them throughout the contract term. The paper presents modeled return scenarios and a proposed put-based hedge against later-period rental-price declines. These are illustrative calculations, not verified realized trading results. [2]

The strategy is most precisely understood as term transformation and physical-capacity intermediation. It is related to market making but not identical to continuously quoting executable two-sided prices. Its profitability depends on occupancy, achieved rental prices, service performance, customer credit, acquisition terms, financing, and working capital throughout the 36-month horizon.

A price hedge addresses only part of that risk. An index put might offset some losses from falling benchmark rental prices if its strike, tenor, quantity, and settlement terms fit the exposure. It does not automatically cover unsold hours, customer defaults, outages, quality differences, geographic or configuration basis, or the risk that the index becomes less representative of the actual inventory. Stress can cause utilization and prices to decline together. Moreover, the proposal does not establish that a hedge of the specified size, tenor, strike, and price is executable today; the paper's own language describes instruments, ratios, and strike levels as contingent on the existence of a liquid market. [2]

> KEY INSIGHT
>
> Hyperbolic's proposal offers a concrete illustration of how an intermediary might assume and manage compute market timing risk commercially. That is different from eliminating it. The trade carries significant duration exposure, operational execution risk, and a hedge that protects one price dimension while leaving utilization, counterparty credit, and basis risks unaddressed. Identifying a timing opportunity and bearing timing risk are the same transaction.

6. A Timing Example: When the Same Demand Produces Different Credit Outcomes

Consider two hypothetical infrastructure projects. Each installs identical hardware and generates $12 million in cumulative customer revenue over three years, but the timing of that revenue differs. Both carry $2 million of annual cash operating costs, $1.5 million of annual scheduled debt service, and an opening unrestricted cash reserve of $1 million. These figures are explicitly illustrative and are not intended to represent current GPU market pricing.

The projects differ only in the arrival of demand:

© October 2026 1337 Advisors. All rights reserved.

$ millions	Year 1	Year 2	Year 3	Three-year total
Project A revenue: early adoption	5.0	4.0	3.0	12.0
Project B revenue: delayed adoption	1.0	4.0	7.0	12.0
Operating cost, each project	2.0	2.0	2.0	6.0
Scheduled debt service, each project	1.5	1.5	1.5	4.5
Project A annual cash after debt service	+1.5	+0.5	−0.5	+1.5
Project B annual cash after debt service	−2.5	+0.5	+3.5	+1.5

Exhibit 2: Illustrative comparison of early and delayed adoption scenarios, showing identical cumulative economics with radically different liquidity paths.

Source: 1337 Advisors. Figures are hypothetical; they do not represent current GPU rental pricing.

Both projects have the same cumulative three-year cash flow after scheduled debt service. Their liquidity paths differ radically. Project A ends Year 1 with $2.5 million in cash, assuming the opening reserve is retained. Project B exhausts its $1 million opening reserve during Year 1 and faces a $1.5 million funding shortfall before the stronger revenues of Years 2 and 3 arrive. Absent a liquidity facility, covenant relief, sponsor support, or restructuring, Project B may not survive to earn its forecast revenue.

This example isolates timing from ultimate demand. It does not imply that delayed projects must fail: reserves, flexible debt structures, customer deposits, equity support, and refinancing can bridge the gap. Nor does it capture a second risk: by Year 3, the installed GPUs may face more efficient substitutes, changing both achievable rental prices and collateral recovery. A project that survives its initial demand gap can still lose economic competitiveness before its debt matures.

> KEY INSIGHT
>
> The financial outcome is determined not only by whether demand eventually materializes, but by whether the asset and its capital structure can survive until it does. That is the economics of timing in its simplest form, and no amount of favorable long-run demand resolves an intermediate-term liquidity gap that the capital structure cannot bridge.

© October 2026 1337 Advisors. All rights reserved.

Part III · The Market That Must Still Be Built

Better price data will not automatically produce better asset valuation. Better asset valuation requires understanding what compute produces, not just what it costs to access.

7. Price Discovery Is Not Value Discovery

Transparent GPU rental quotations and standardized contracts can reduce transaction friction. They can improve procurement, price comparison, and certain forms of hedging. But a GPU-hour index prices a defined access service, not the economic value of the output produced by every configuration, and not the recoverable value of every asset financed against that service.

For a buyer, the relevant comparison is the quality-adjusted cost of completing a workload. For an operator, it is the future earning power of the installed fleet after operating costs, maintenance, upgrades, and utilization. For a lender, it is the distribution of recoveries under plausible defaults and redeployment conditions. For a physical intermediary, it is the distribution of net cash flows after acquisition commitments, sales realization, operating costs, and hedge basis. The same rental index feeds each of these analyses without answering any of them on its own.

The compute forward curve will also require conceptual discipline. A forecast of future spot rental rates is not an executable forward curve. A tradable forward price reflects the specified hardware or service, delivery location, quality and availability, tenor, counterparty and delivery risk, financing, liquidity, and the compensation required to warehouse uncertainty. A physically delivered capacity right is not interchangeable with a financially settled benchmark future simply because both reference a GPU-hour. This is where the instrument-design work in WP-02 matters: transferability, precise delivery specifications, remedies, and enforceability are not administrative details. They determine what economic exposure can actually be bought, sold, financed, or hedged. [3]

Three forces can pull compute economics in different directions simultaneously, and their relative timing determines prices by vintage, geography, workload, and tenor. Productivity improvements reduce machine-hours required per unit of useful output, arriving on model-release cadences measured in months to years. Adoption expansion increases total demand for useful output,

© October 2026 1337 Advisors. All rights reserved.

moving on enterprise and application cycles measured in years. Physical constraints limit where and when productive capacity can be delivered, shaped by power and interconnection lead times measured in years to decades. None of these forces synchronizes with the others.

8. Why Intelligence Demand Need Not Become GPU Scarcity

The intuitive inference that more valuable intelligence must make every GPU more valuable is not economically valid. Demand for useful output can rise while the machine-hours needed to produce each unit fall. Whether total capacity demand increases depends on the relative pace of useful-output demand and production efficiency, as well as substitution among hardware types and locations.

There are plausible scenarios in both directions. Better models and hardware could sharply reduce the compute required for existing tasks, weakening rental rates for particular vintages even as aggregate AI use grows. Lower production costs could also unlock applications that were previously uneconomic, causing aggregate demand to expand faster than efficiency improves. Demand may grow in one workload category and contract in another. Frontier training, latency-sensitive inference, batch inference, and simulation need not share a single demand or supply curve.

Strong current contracting and uncertainty about future utilization are not mutually exclusive. A sold-out cluster today can coexist with uncertain demand for the next generation of capacity. Conversely, an underutilized facility in one region does not establish that the entire industry is overbuilt. The relevant analysis is granular: which workload, which hardware, which geography, which contract tenor, and at what point in the financing cycle?

A useful framework must remain open to evidence that reverses its initial expectations. If demand expands faster than productivity improves, scarcity can intensify. If productivity outpaces economically viable use, some installed capacity loses value. The thesis is about the consequences of their relative timing, not a predetermined direction for either.

© October 2026 1337 Advisors. All rights reserved.

9. Where Scarcity May Migrate

Scarcity is not permanently attached to the GPU. At different moments, the binding constraint may be accelerator availability, high-bandwidth interconnect, electricity, cooling, suitable sites, trained operators, customer integration, or financeable credit. Different constraints create rents for different assets.

An accelerator generation can lose relative earning power while a facility with scarce, reliable power and suitable infrastructure retains value. But even that relationship is conditional: workloads may migrate, alternative power-secured sites may come online, or efficiency improvements may reduce the need for some physical capacity. Latency, sovereignty, networking, and operational continuity can anchor certain workloads geographically while others remain highly mobile.

The market should therefore ask not only what an asset is worth today, but which constraint will matter at the point when its cash flow must be earned, refinanced, or recovered. For credit investors, that is a residual-value question. For operators, it is a reinvestment question. For intermediaries, it is a relative-value and term-structure question.

> KEY INSIGHT
>
> The question lenders are beginning to ask, what is the hardware worth if we get it back, is the right question at the wrong level of specificity. The answer depends on which dimension of scarcity will be binding at the time of recovery, what workloads will still be viable on that hardware generation, and whether the facility can attract the demand that makes it worth operating. A rental-rate index establishes price discovery. A model of scarcity migration answers the harder question.

10. The Measurement Infrastructure: From Benchmarks to Bankable Economics

The industry is not starting from zero. Workload benchmarking efforts, including MLPerf, already provide structured ways to compare performance under defined conditions. Providers, brokers, and market venues publish various forms of compute pricing. These are meaningful foundations, but they answer different questions and should not be conflated. [4]

© October 2026 1337 Advisors. All rights reserved.

The missing integration is between reproducible performance, commercially realizable pricing, all-in production economics, and asset-level financial risk. A practical framework would use defined workload baskets with explicit model, quality, latency, throughput, memory, and reliability requirements. It would compare feasible hardware and topology configurations on effective useful output, energy consumption, operating cost, and achievable utilization. Those results would be joined, with clear provenance, to prices that distinguish advertised offers, executable quotes, and completed transactions.

The next layer would estimate future earning power under alternative adoption, technology, price, and utilization scenarios. Residual valuation should distinguish continued operation, redeployment, and liquidation. Financing analysis should connect those scenarios to covenants, amortization, customer concentration, debt maturity, and the availability of sponsor liquidity. Hedge analysis should test basis and joint stress, including the possibility that utilization, rental prices, and customer credit weaken simultaneously.

The objective is not a universal dollar value for a unit of intelligence. No single measure can equate the value of a trading prediction, a medical workflow, and a software-generation task. The achievable objective is narrower and more useful: comparable workload production economics, transparent uncertainty, and better estimates of the risks actually embedded in financing and intermediation contracts.

1337 Advisors: Research and measurement agenda

The 1337 Advisors platform research program separates access price from workload-adjusted cost, technological from economic obsolescence, and going-concern value from recovery value. [5] It is directed toward connecting provider pricing, benchmark results, and configuration data to those questions.

The underlying data require careful classification. A posted price is not necessarily an executable quote; an executable quote is not necessarily a completed transaction. A benchmark for a GPU model is not automatically representative of a rented cluster with different topology, memory, software, or service terms. Provider counts, observation counts, benchmark joins, and topology coverage are meaningful only alongside a dated, reproducible methodology identifying their sources, coverage, matching rules, and limitations. This paper does not present platform data as proof of transaction-price discovery or workload-level valuation accuracy.

That restraint matters. A more sophisticated measurement system may not make every project appear safer. It may reveal that some assets depend on optimistic utilization assumptions, fragile contractual rights, or recovery values that have never been tested against a stressed market. Its

© October 2026 1337 Advisors. All rights reserved.

value is in distinguishing those exposures, not in manufacturing reassuring marks.

Conclusion

The compute market is not facing one question about whether AI demand will be large enough. It faces several related questions: when useful applications become economically viable, when productive capacity becomes available, when technological change alters the cost of production, when scarcity shifts among inputs, and when financing obligations must be serviced or repaid. Those events do not occur on a common schedule.

Stanley Lee's credit analysis illustrates why lenders can become more discriminating even when underlying demand is real: the capital structures supporting infrastructure projects must survive the interval between commitment and payoff, and the tools to value, transfer, or hedge the associated risks are still being built. Hyperbolic's term-transformation proposal illustrates how an intermediary may earn a premium by connecting suppliers and customers operating on different contractual horizons, while accepting the risks that bridge creates. Neither advances the overarching timing thesis. Their relevance is that both expose financial consequences of the same underlying economic problem.

The manufacturing lens provides the connection. It distinguishes the machine from the factory; the output of computation from useful intelligence; useful intelligence from the value its consumer realizes; and the price of access from the earning and recovery capacity of the underlying asset. Once those distinctions are made, timing becomes a variable that can be analyzed rather than a narrative attached after the fact.

WP-01 argued for the reallocation of future compute access. WP-02 explored the design of a transferable Future Compute Right. This paper adds a condition for making such a market durable: the right cannot be valued, financed, or hedged intelligently without understanding what the underlying capacity can produce, for whom, at what cost, and over what economic life.

The economics of intelligence will not be determined simply by how much compute is built or how capable the technology becomes. They will also be determined by when customers can realize value from that capability, and whether the assets and capital structures built to supply it can endure the interval.

Sources, attribution, and analytical boundaries

© October 2026 1337 Advisors. All rights reserved.

[1] Stanley Lee, "The compute credit cycle just turned, and why that's a good thing," Liquid Compute (article provided to 1337 Advisors). His analysis concerns the differentiation of compute credit, operator risk, collateral, concentration, and market plumbing. He does not advance the overarching timing thesis of this paper. Numerical market anecdotes in his article have not been independently verified here.

[2] Hyperbolic Labs, *Compute Market-Making White Paper: Buying long-dated capacity, selling short duration, and hedging the tail* (September 2026; supplied to 1337 Advisors). Transaction assumptions, modeled returns, and hedging illustrations are proposals, not evidence of executed trades or executable option markets. The paper's own language identifies instruments, ratios, and strike levels as subject to determination at execution and contingent on the existence of a liquid market.

[3] 1337 Advisors, *WP-01* and *WP-02*, published research. References summarize their conceptual role in this series and do not imply that the proposed instrument is currently liquid or widely adopted.

[4] MLCommons, MLPerf benchmark initiatives. Benchmark results are useful evidence about performance under defined conditions; they are not, on their own, verified commercial transaction prices or customer economic returns. mlcommons.org/benchmarks

[5] 1337 Advisors Platform, platform.1337advisors.com. Covers GPU pricing across multiple providers, MLPerf-linked benchmark results, and provider configuration data. Methodology and coverage details are available through the platform's research documentation.

Scope: This is conceptual economic research, not investment advice, a market-wide overbuild or underbuild determination, a claim about undisclosed strategies or returns at any named firm, or a verified empirical forecast. The numerical exhibit in Section 6 is explicitly hypothetical. Statements about emerging market structure are analytical characterizations, not assertions that every venue, instrument, or bilateral trade described is currently executable.

© 2026 1337 Advisors. All rights reserved.