Pricing Negentropy: A Quotation Model for Pre-Processed Context Bundles in Machine-to-Machine Knowledge Markets

Nicolas Limare — ag3ntlab *Version 1.0 — July 2026*

Abstract

AI agents acquire task-relevant information in one of two ways: by burning inference tokens to search, crawl, filter, and synthesize raw sources, or — as machine-to-machine (M2M) payment rails such as x402 mature — by purchasing pre-processed, directly consumable context from another agent. The first path pays, at every use, the full cost of reducing informational entropy; the second amortizes that cost across many buyers. We propose, to our knowledge, the first explicit quotation model for such context bundles. The model prices a bundle by the processing cost it saves its buyer — a function of the bundle's *refinement ratio* (raw tokens an agent would need to ingest per curated token delivered) — bounded above by the buyer's decision-theoretic value of information, discounted by an information-diffusion decay, and split by bargaining weight. Calibrated against July-2026 inference prices and measured deep-research agent costs, the model predicts one-shot bundle prices of \$0.5–\$4, independently consistent with observed x402 micropayment sizes (\$0.001–\$5, mean ≈ \$0.20). We show that the production side admits a systematic temporal arbitrage — refining at batch prices (0.5×) what buyers would otherwise compute at real-time or priority prices (1×–2.5×) — and we state the competitive conditions under which any such price survives. A deterministic reference implementation of the quoting algorithm accompanies the paper.

Keywords: value of information, rational inattention, data markets, AI agents, micropayments, x402, knowledge curation.

1. Introduction

Large-language-model agents now perform a substantial share of information work: they search, read, deduplicate, verify, and synthesize. Each of these operations is metered — per-token inference pricing makes the cost of *reducing uncertainty* explicit in dollars for the first time in economic history. Simultaneously, payment rails designed for machine-to-machine transactions have become production infrastructure: the x402 protocol (HTTP 402 with stablecoin settlement) moved under the Linux Foundation in 2026 with backing from major payment networks and cloud providers [20], and content-delivery infrastructure is repricing information access around *actual use in an AI answer* rather than crawling [22].

This creates a textbook make-or-buy decision, executed autonomously by software: an agent that needs context can either (i) assemble it from raw sources, paying inference and tool costs plus latency, or (ii) buy a curated, structured, immediately consumable *bundle* from a specialist. What is missing is a pricing law. Recent work on agentic marketplaces provides simulation environments and mechanism scaffolding [16, 17, 18], but — as far as we could establish by systematic search — no closed-form quotation model for pre-processed context sold to agents. This short paper proposes one.

Our contributions are: (1) a buyer-side willingness-to-pay bound built on measurable quantities (bundle size, refinement ratio, live token prices) with an explicit information-theoretic skeleton; (2) the corrections classical theory imposes on the naive bound — a decision-theoretic cap, a bargaining split, an information-diffusion decay, and a differentiation condition without which the price collapses; (3) a 2026 calibration against measured deep-research agent costs, which lands, independently, in the empirically observed x402 micropayment range; and (4) a deterministic reference implementation.

2. Related work

Value of information. The decision-theoretic value of information originates with Raiffa and Schlaifer's EVPI/EVSI [2] and Howard's information value theory [1]. Two structural results matter here: VoI is non-negative for a Bayesian expected-utility maximizer, and VoI is zero whenever a signal cannot change the optimal action — it is a property of the *decision*, not of the bit count. Blackwell's ordering [3] formalizes why mutual information alone can misrank two signals of equal entropy. Bergemann, Bonatti and Smolin [13] characterize the revenue-

maximizing sale of information as a menu of statistical experiments priced by the buyer's *incremental* value over what she already knows — the closest analytic precedent to our ΔI-driven term.

Information economics. Stigler [4] gives the optimal-search stopping rule that defines our buyer's outside option; Arrow [5] states the fundamental disclosure paradox — the buyer cannot value information without acquiring it — which any real bundle market must engineer around; Shapiro and Varian [6] establish the cost structure (high first-copy cost, near-zero marginal cost) that forces value-based rather than cost-based pricing, and the versioning strategies that follow; Grossman and Stiglitz [7] guarantee an equilibrium rent to costly information aggregation *and* imply that the rent erodes as the information diffuses — the mechanism behind our decay term.

Rational inattention. Sims [8] models attention as a Shannon channel of limited capacity; the Lagrange multiplier λ on mutual information is a *shadow price of information* in utility per nat — the formal bits-to-dollars bridge this paper relies on. Matějka and McKay [9] show the resulting optimal choice rule is multinomial logit with λ as temperature. In our setting, λ is not a psychological constant but a *market observable*: the metered cost at which the buying agent can process tokens.

Thermodynamics of computation. Szilard [10], Brillouin [11], Landauer [12], and Bennett [15] establish the physical floor: acquiring or erasing one bit costs at least $kT \ln 2 \approx 2.87 \times 10^{-21}$ J at 300 K. We use this only to locate the market: the *economic* cost of extracting one useful bit from the raw web with 2026 LLMs exceeds the physical floor by roughly twenty-five orders of magnitude (§5.3). Information sorting is the least physically efficient large-scale operation in the economy; the gap is the space in which a curation market lives.

Data markets. Agarwal, Dahleh and Sarkar [14] price training data by Shapley value with replication-robustness for non-rival goods; Bergemann, Bonatti and Gan [19] analyze data intermediation under non-rivalry and externalities. The 2024–2026 agent-market literature [16, 17, 18] contributes environments and empirical protocols, not pricing laws — the gap this paper addresses.

3. The model

3.1 Setup: information density and the refinement ratio

A buyer agent faces a task requiring it to move its knowledge state from prior uncertainty H_0 to a target H*; write $\Delta I = H_0 - H$* for the task-relevant information to acquire (bits). This information exists, diluted, in raw sources. Define:

- **d_raw** — useful-bit density of raw sources (task-relevant bits per token ingested when crawling/reading);
- **d_B** — useful-bit density of the curated bundle B (bits per bundle token);
- **ρ = d_B / d_raw** — the **refinement ratio**: raw tokens an agent must process per curated token of equal informational contribution.

ρ is the central, *measurable* quantity of the model: a seller who logs production (tokens ingested per token shipped) observes it directly, and a buyer can estimate it from the token telemetry of its own research runs. Measured deep-research agent runs in 2026 imply ρ ≈ 5–25 (§5.1), with published figures being lower bounds because reasoning tokens and report tokens are bundled in provider telemetry.

A bundle of size S tokens therefore delivers ΔI = d_B · S bits, which the buyer would otherwise have had to extract from ρ·S raw tokens.

3.2 The buyer's make-or-buy problem

Let c_in, c_out be the buyer's real-time per-token prices (input, output); c_tool·n its per-run tool spend (search/crawl APIs); σ the synthesis overhead (intermediate notes written and re-read, empirically ≈ 2–3); w·Δτ the buyer's valuation of the latency difference between do-it-yourself research (5–30 minutes measured) and near-instant bundle retrieval; r_fail a risk premium for failed or incomplete assembly (agentic run costs vary up to 30× run-to-run on identical tasks [23]); and C_verif the buyer's cost of verifying/trusting the bundle.

Do-it-yourself: C_DIY = (1 + r_fail) · [c_in·ρ·S + c_out·σ·S + c_tool·n] + w·τ_DIY

Buy: purchasing does not exempt the buyer from *reading* the bundle: C_BUY = P + c_in·S + C_verif + w·τ_B

Willingness-to-pay bound (buyer indifference, Δτ = τ_DIY − τ_B):

P_max = (1 + r_fail) · [c_in·ρ·S + c_out·σ·S + c_tool·n] + w·Δτ − c_in·S − C_verif (1)

3.3 The information-theoretic skeleton

Substituting S = ΔI/d_B and ρ = d_B/d_raw into the dominant term of (1):

c_in·S·(ρ − 1) = c_in · ΔI · (1/d_raw − 1/d_B) (2)

The price of a bundle is proportional to its task-relevant information content times the *difference of inverse useful densities* between raw sources and the curated artifact. Since 1/d is tokens-per-bit and c_in is dollars-per-token, c_in/d is the dollar cost of acquiring one bit at a given density: equation (2) is a *difference of per-bit acquisition costs*, i.e. exactly the linear avoided-cost form that rational inattention licenses (the buyer's λ, in this market, *is* its metered processing cost). All of the seller's craft lives in two terms: raising d_B (organization, structure, deduplication, machine-readability) and lowering C_verif (provenance, citations, verifiable packaging).

3.4 The value-side cap

Equation (1) is a *substitution* bound: the buyer will not pay more than the cost of making the bundle itself. Decision theory imposes a second, independent cap: the buyer will not pay more than the information is worth to its decision, ΔEVSI — its expected value of sample information *incremental to what it already knows* [1, 2, 13]. Crucially, ΔEVSI is not monotone in Shannon content: bits that never change the buyer's action are worthless regardless of their entropy [3]. The operative ceiling is therefore

P_ceiling = min[ΔEVSI , P_max] (3)

For commodity-grade context (facts available elsewhere), P_max binds; for exclusive, decision-critical information, ΔEVSI binds. A rational seller segments accordingly.

3.5 Temporal decay

Information rents erode as the information diffuses [7]. We adopt a half-life parameterization: a bundle constituted at time 0 and sold at age t quotes

P(t) = P(0) · 2^(−t/t½) (4)

with t½ set by the domain (hours for news; weeks–months for technical documentation; years for stable theory). The strategic consequence: **continuously maintained bundles (age ≈ 0 at every sale) are the only durably priced good**; static bundles decay toward zero. Maintenance, not constitution, is the moat.

3.6 The seller: amortization and the refinery spread

The seller produces once and sells N times (information is non-rival; delivery is ≈ free). Production can be scheduled at the cheapest available compute: all three major inference providers offer batch processing at 0.5× real-time prices, while buyers who need answers now pay 1× (standard) to 2–2.5× (priority tiers) — a measured, cross-provider immediacy spread of up to 5:1 (§5.1). Production cost:

$C_{prod} = 0.5 \cdot [\, c_{in} \cdot \rho \cdot S + c_{out} \cdot \sigma \cdot S \,] + c_{tool} \cdot n$ (batch, off-peak, gathering on cheaper models)

Profitability requires only $P \geq C_{prod}/\hat{N} + c_{delivery}$ for expected sales $\hat{N}$. The seller operates an *information refinery with a warehouse*: it buys compute forward at the stored-commodity price, transforms it into negentropy whose only carrying cost is the decay $\lambda = \ln 2/t_{½}$ of (4), and sells immediacy at spot. Four independent multipliers compose the margin: the batch spread (2×), the immediacy premium (up to 2.5×), amortization over N buyers, and the refinement ratio ρ itself (each bundle token spares the buyer ρ−1 raw tokens).

3.7 Equilibrium: the quotation formula

Between the seller's floor and the buyer's ceiling, where the price lands is a bargaining and competition question.

Bilateral trade. With marginal cost ≈ 0, the Nash bargaining solution [21] gives $P^* = \beta \cdot P_{ceiling}$ with β the seller's share of surplus (β = ½ under symmetry; higher when the buyer has no alternative supplier — the current, empty-market case; lower under competition).

Competition. Two sellers of substitutable bundles are Bertrand competitors in a zero-marginal-cost good: the equilibrium price is zero *regardless of ΔI*. A quoted price survives only on differentiation that a rival cannot cheaply replicate: freshness under continuous maintenance (§3.5), source exclusivity, or verifiable trust (low C_{verif}). This is not a defect of the model but its sharpest prediction: *undifferentiated curation is worthless in equilibrium; maintained, verifiable curation is not.*

Transparency. Uniquely among markets, both sides can *compute* the outside option: a buying agent can estimate its own C_DIY in a few hundred tokens, and the seller can estimate the marginal buyer's. Willingness-to-pay is algorithmically observable, so quotation becomes a deterministic function rather than a negotiation — a property with no clean analogue in human markets.

Assembling (1), (3), (4):

$P^*(B, t) = \beta \cdot \min[\, \Delta EVSI\,, P_{max}(B)\,] \cdot 2^{(-t/t_{1/2})}$, subject to $P^* \geq C_{prod}/\hat{N}$ (5)

Arrow's paradox and its mitigation. The realized price is further capped by the buyer's *expected* (pre-verification) value [5]: information cannot be inspected without being acquired. Structured knowledge formats mitigate this structurally: a bundle whose index (table of contents, coverage map, provenance, checksums) is free to read while its content is paid separates *evaluation* from *acquisition* — the index is a verifiable sample that does not exhaust the good. Reputation and reproducible production records close the remaining gap; every dollar of C_verif so removed reappears in P_max at every sale.

4. Calibration (July 2026)

Inputs (sourced, first-party pricing pages except where noted). Real-time input prices $c_{in} \approx$ \$2–5 per Mtok for frontier-adjacent models; batch at 0.5× uniformly across the three major providers; priority tiers at 2.0–2.5×; cached-context reads at 0.1× — notably, the providers themselves price *already-processed context* at a 90% discount, an independent confirmation of the density economics. Measured deep-research runs: \$0.4–\$8 typical total cost (extremes to ~\$30), 5–30 minutes latency, with 30× run-to-run cost variance [23]; derived refinement ratios $\rho \approx 2.5$–22 (lower bounds; honest range 5–25). Tool spend \$0.1–\$1 per run at current search-API prices (\$0.005–\$0.016 per query).

Worked example. A deep-research-grade bundle, S = 20k tokens, $\rho = 15$, $\sigma = 2.5$, $r_{fail} = 0.3$, c_{in} = \$3/Mtok, c_{out} = \$15/Mtok, tools \$0.30, immediacy premium \$0.50, C_{verif} = \$0.05, $\beta = ½$, age 10 days, $t_{½}$ = 90 days: equation (5) quotes **$P^* \approx$ \$0.9** (sensitivity $\rho \in [5, 25]$: \$0.7–\$1.1). Across plausible bundle sizes and freshness, the model concentrates one-shot quotes in **\$0.5–\$4**.

Independent consistency check. The x402 rail — designed with no reference to this model — transacts micropayments in the \$0.001–\$5 range with an observed mean ≈ \$0.20 [24]. Theory and rail land on the same interval without communicating. A continuously

maintained bundle consumed regularly by one agent quotes, at the same parameters, ≈ \$5–20/month in repeated pulls.

5. The market today

5.1 The rail exists; the category is empty. x402 settles in USDC with facilitator fees of \$0.001/tx [25]; discovery layers exist (service bazaars with agent-facing search). Yet a systematic search (July 2026) finds *no named provider selling pre-processed context per-request with a public price sheet* — the nearest neighbors are per-crawl/per-use content licensing at the CDN layer [22] and raw-dataset marketplaces. The category this paper prices is, at writing, unoccupied.

5.2 Honest sizing. Independent on-chain analysis measures real x402 volume at roughly \$28k/day with ~50% artificial activity [24]. The addressable demand today is small; the model's near-term value is positional and infrastructural, not fiscal.

5.3 The economic Landauer gap. Extracting one useful bit from raw web sources at 2026 prices costs of order 10^{-4}–10^{-3} dollars (c_in/d_raw); the physical floor is kT ln2 ≈ 2.87×10^{-21} J ≈ 10^{-29} dollars of electricity. The ~10^{25} gap measures how far information sorting sits from its thermodynamic limit — and hence how much room the refining market has to exist, compete, and compress.

6. Reference implementation

A deterministic quoting tool accompanies this paper: given a knowledge bundle (structured markdown with machine-readable index and provenance), it audits conformance and freshness, measures S, applies equation (5) under a declared parameter set, and emits a quote with per-term breakdown and sensitivity bounds. No model inference occurs in the quoting or selling path — pricing is a function, not an agent, by design (fixed-price settlement is native to the rail; a negotiating LLM would add cost and an injection surface without a protocol slot to negotiate in). *[Live store URL — to be inserted at publication.]*

7. Limitations

(i) ΔI and the densities d are latent; the operational model runs entirely on proxies (S, ρ, logged production telemetry), and the information-theoretic form (2) is the skeleton that justifies their structure, not a measurement claim. (ii) $\Delta EVSI$ is buyer- and task-specific; sellers observe it only through segmentation and demand. (iii) ρ estimates inherit provider telemetry that bundles reasoning with reporting tokens; published values are lower bounds. (iv) The buyer-rationality premise (agents that compute C_{DIY} before buying) is native to agentic buyers but empirically young; observed demand on the rail remains small and partially artificial [24]. (v) Half-life decay is a parameterization of [7], not a derived law; no canonical VoI-decay theorem exists. (vi) The model prices *bundles*, not streams; subscription pricing here is repeated per-pull settlement of a maintained bundle, which the rail supports natively.

8. Conclusion

When both the cost of reducing uncertainty and the payment for its product are metered in the same unit, the price of organized information stops being mystical: it is the buyer's avoided processing, capped by decision value, decayed by diffusion, split by bargaining power, and defended only by maintenance and verifiable trust. The formula prices the work of Maxwell's demon at market rates — some twenty-five orders of magnitude above the thermodynamic floor the demon actually owes.

References

[1] R. A. Howard, "Information Value Theory," *IEEE Transactions on Systems Science and Cybernetics* 2(1), 22–26, 1966.

[2] H. Raiffa, R. Schlaifer, *Applied Statistical Decision Theory*, Harvard University, 1961.

[3] D. Blackwell, "Equivalent Comparisons of Experiments," *Annals of Mathematical Statistics* 24(2), 265–272, 1953.

[4] G. J. Stigler, "The Economics of Information," *Journal of Political Economy* 69(3), 213–225, 1961.

[5] K. J. Arrow, "Economic Welfare and the Allocation of Resources for Invention," in *The Rate and Direction of Inventive Activity*, NBER/Princeton, 609–626, 1962.

[6] C. Shapiro, H. R. Varian, *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business School Press, 1999.

[7] S. J. Grossman, J. E. Stiglitz, "On the Impossibility of Informationally Efficient Markets," *American Economic Review* 70(3), 393–408, 1980.

[8] C. A. Sims, "Implications of Rational Inattention," *Journal of Monetary Economics* 50(3), 665–690, 2003.

[9] F. Matějka, A. McKay, "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model," *American Economic Review* 105(1), 272–298, 2015.

[10] L. Szilard, "Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen," *Zeitschrift für Physik* 53, 840–856, 1929.

[11] L. Brillouin, "The Negentropy Principle of Information," *Journal of Applied Physics* 24(9), 1152–1163, 1953.

[12] R. Landauer, "Irreversibility and Heat Generation in the Computing Process," *IBM Journal of Research and Development* 5(3), 183–191, 1961.

[13] D. Bergemann, A. Bonatti, A. Smolin, "The Design and Price of Information," *American Economic Review* 108(1), 1–48, 2018.

[14] A. Agarwal, M. Dahleh, T. Sarkar, "A Marketplace for Data: An Algorithmic Solution," *Proceedings of the 2019 ACM Conference on Economics and Computation (EC '19)*, 701–726, 2019.

[15] C. H. Bennett, "The Thermodynamics of Computation — a Review," *International Journal of Theoretical Physics* 21(12), 905–940, 1982.

[16] T. Sashihara et al., "LLM-based Multi-Agent System for Simulating Strategic and Goal-Oriented Data Marketplaces," arXiv:2511.13233, 2025.

[17] "Magentic Marketplace: An Open-Source Environment for Studying Agentic Markets," arXiv:2510.25779, 2025.

[18] "Agent Bazaar: Enabling Economic Alignment in Multi-Agent Marketplaces," arXiv:2605.17698, 2026.

[19] D. Bergemann, A. Bonatti, T. Gan, "The Economics of Social Data," *RAND Journal of Economics* 53(2), 263–296, 2022.

[20] Linux Foundation, "Launching the x402 Foundation," press release, 2026.

[21] J. F. Nash, "The Bargaining Problem," *Econometrica* 18(2), 155–162, 1950.

[22] Cloudflare, "Introducing Pay Per Crawl," July 2025; and the July 2026 transition to per-use compensation ("pay per use").

[23] Stanford Digital Economy Lab, "How are AI agents spending your tokens?", May 2026.

[24] CoinDesk / Artemis, "Coinbase-backed AI payments protocol wants to fix micropayments, but demand is just not there yet," March 2026.

[25] Coinbase Developer Platform, x402 documentation (facilitator pricing), accessed July 2026.