CCIR WORKING PAPER NO. 2026-01

From Rental Rates to Residual Value:

A Two-Stage Framework for GPU Collateral Analytics

Compute Credit Index Research (CCIR)

April 2026

research@ccir.io
https://github.com/ccir-index

ABSTRACT

The GPU-backed lending market exceeds $20 billion in outstanding obligations, yet participants lack the standardized analytical infrastructure available in every other mature asset-backed market. Depreciation assumptions vary by a factor of six across market participants. No independently verifiable rental rate reference exists that meets the governance standards required for credit document citation. No market-derived depreciation curves have been published. This paper proposes a two-stage framework: Stage 1 establishes a transparent, open-methodology rental rate index (CRI-H100) providing the revenue-side input to collateral analysis; Stage 2 derives market-based depreciation curves from cross-generational rental rate time series, replacing ad hoc management estimates. We draw on commodity financialization theory, electricity market design (locational marginal pricing), and equipment-backed securitization precedents (aircraft EETCs, railcar ABS) to argue that the sequencing is structurally necessary and that the credit-specific analytical layer — distinct from existing derivatives-oriented indices — is the binding constraint on market development. We present the index methodology, a statistical validation framework, preliminary empirical observations from the initial data collection period, and specific credit market applications including covenant design, ABS trigger structures, and rating agency interface specifications.

Keywords: GPU compute, asset-backed securities, collateral valuation, depreciation, benchmark methodology, reference rate, open methodology

JEL Classification: G12, G23, G28, L86

Companion documents: CCIR Methodology v1.1.0 • CCIR Governance Framework v1.0

1. Introduction

Five American technology companies are projected to spend over $700 billion in capital expenditure in 2026, primarily on data center infrastructure for artificial intelligence (The Economist, 2026). By comparison, the global oil and gas industry spends approximately $500 billion annually on exploration and production. GPU compute infrastructure has become one of the largest categories of capital deployment in economic history.

Yet the financial infrastructure surrounding this asset class remains primitive. GPU-backed loans carry effective interest rates of approximately 14% on non-investment-grade tranches — roughly triple the rate on investment-grade corporate debt — reflecting in part the lender's inability to model collateral risk with precision (CoreWeave, Inc., Form S-1, 2025). The largest transactions are not genuinely asset-backed but credit-substituted, underwritten against hyperscaler balance sheets rather than the GPU hardware itself: CoreWeave's most recent $8.5 billion facility is expected to receive an A-minus rating based on Meta's credit profile, not the underlying GPU collateral (Bloomberg, 2026). A meaningful portion of institutional lenders have opted out entirely, unable to develop defensible collateral valuation methodologies (PitchBook, 2025).

This paper addresses a specific and remediable gap: the absence of standardized analytical infrastructure for GPU collateral valuation in credit markets. We propose a two-stage framework in which a transparent rental rate reference index is established first (Stage 1), and market-derived depreciation curves are constructed from cross-generational rental rate data second (Stage 2). The sequencing is structurally necessary: credible depreciation curves require a standardized price signal as their foundation.

Our framework draws on three bodies of literature that have not been connected to GPU compute markets:

First, commodity financialization theory (Gorton and Rouwenhorst, 2006; Tang and Xiong, 2012) identifies the conditions under which physical assets develop financial market infrastructure. We apply these conditions diagnostically to GPU compute and identify which conditions are met, which are absent, and which require novel solutions.

Second, electricity market design — specifically locational marginal pricing (Joskow and Tirole, 2000; Hogan, 1992) — provides the closest structural analogue to compute pricing. Like electricity, GPU compute is economically impractical to store at scale, geographically fragmented, and subject to real-time supply-demand dynamics that create persistent regional price divergence. We argue that the LMP framework, suitably adapted, offers the correct analytical basis for a compute reference rate.

Third, equipment-backed securitization precedents — aircraft Enhanced Equipment Trust Certificates (Littlejohns and McGairl, 2006), railcar ABS, and semiconductor equipment financing — demonstrate how depreciating technology collateral has been successfully integrated into structured credit markets through standardized appraisal methodologies, depreciation conventions, and stress-testing frameworks.

The paper proceeds as follows. Section 2 reviews the relevant literature and establishes the theoretical framework. Section 3 presents the current state of the GPU-backed lending market, including a systematic analysis of existing depreciation assumptions and their implications. Section 4 introduces the two-curve framework. Sections 5 and 6 detail the Stage 1 and Stage 2 methodologies, respectively. Section 7 presents the statistical validation framework. Section 8 describes credit market applications. Section 9 addresses intra-model performance variance and its implications for collateral heterogeneity. Section 10 discusses limitations and areas for further research. Section 11 concludes.

2. Literature Review and Theoretical Framework

2.1 Commodity Financialization: When Do Physical Assets Develop Financial Infrastructure?

The foundational question underlying this paper is: what conditions must obtain for a physical asset to develop the financial market infrastructure — benchmarks, derivatives, securitization vehicles — that enables efficient capital allocation? The commodity financialization literature identifies several necessary conditions.

Gorton and Rouwenhorst (2006) establish that commodity futures returns are driven by the convenience yield, scarcity, and inventory dynamics of the underlying physical asset. For financial infrastructure to develop, the asset must be (a) standardizable, (b) storable or continuously produced, (c) subject to price risk that market participants have economic incentive to hedge, and (d) measurable with sufficient transparency for counterparties to agree on a price.

Tang and Xiong (2012) document how the financialization of commodity markets after 2004 increased price correlations across previously independent commodity classes, suggesting that the act of creating financial infrastructure around a commodity changes the asset's price dynamics in ways that must be anticipated in benchmark design.

Applying these conditions to GPU compute: condition (a) is partially met — GPU models are standardized by manufacturer specification, but intra-model performance variance (discussed in Section 9) complicates fungibility. Condition (b) is met for rental compute — GPU-hours are continuously produced and effectively perishable (unused capacity represents lost revenue that cannot be recovered). Condition (c) is strongly met — AI companies, data center operators, and lenders all face material GPU price risk. Condition (d) is the binding constraint — there is no agreed transparent benchmark, and existing price discovery mechanisms are fragmented across proprietary platforms.

Our contribution addresses condition (d) directly, while disclosing the limitations imposed by the partial satisfaction of condition (a).

2.2 Electricity Market Design: Locational Marginal Pricing as Structural Analogue

GPU compute shares critical structural properties with electricity that make electricity market design the most instructive precedent for compute pricing:

Non-storability and perishability. Like electricity, an idle GPU-hour represents revenue permanently lost — it cannot be economically stored for later sale. This creates real-time supply-demand dynamics that produce price volatility and regional divergence — dynamics substantially muted in storable commodities like oil where inventory buffers absorb short-term demand fluctuations.

Geographic fragmentation. Hogan (1992) demonstrated that electricity prices vary by geographic node due to transmission constraints. GPU compute exhibits analogous geographic fragmentation: rental rates differ across US regions and between countries, driven by power costs, cooling infrastructure, data sovereignty requirements, and latency constraints. Unlike oil, compute cannot be “shuttled around” at low marginal cost.

Capacity constraints and congestion pricing. Joskow and Tirole (2000) analyze how transmission congestion produces locational price divergence in electricity markets. GPU compute faces analogous capacity constraints: specific GPU architectures in specific regions command price premiums during periods of high demand, and the interconnect bandwidth between data centers creates functional “transmission constraints” for distributed workloads.

The LMP framework resolves these dynamics by computing nodal prices that reflect both the marginal cost of generation and the congestion cost of delivery. We argue that a compute reference rate must similarly account for geographic price structure — not by computing locational marginal prices (which would require network topology data unavailable in current GPU markets), but by publishing regionally segmented indices that make the geographic price structure transparent and measurable.

The California electricity crisis of 2000–2001 provides a cautionary precedent: financialization of an infrastructure commodity without adequate market design and price transparency produced catastrophic outcomes. The lesson is not that financial infrastructure is dangerous, but that it must be constructed with transparency and governance commensurate with the complexity of the underlying market (Borenstein, Bushnell, and Wolak, 2002).

2.3 Equipment-Backed Securitization: Depreciating Technology Collateral in Credit Markets

The structured finance literature offers direct precedents for financing depreciating technology assets.

2.3.1 Aircraft Enhanced Equipment Trust Certificates (EETCs)

Aircraft EETCs are the most developed securitization market for high-value, depreciating physical assets. When Northwest Airlines issued the first EETC in 1994, senior tranches priced at approximately 300 basis points over Treasuries, reflecting uncertainty about a novel structure (Littlejohns and McGairl, 2006). As the market developed standardized appraisal methodologies, depreciation assumptions validated through multiple credit cycles, and a track record of recoveries, spreads compressed to 70–150 basis points for creditworthy issuers.

The aircraft precedent is instructive in three specific ways. First, aircraft depreciate — but the rate of depreciation depends on model, configuration, maintenance history, and market conditions, creating collateral valuation challenges directly analogous to GPUs. Second, aircraft appraisal became standardized through independent appraisal firms (AVITAS, Morten Beyer & Agnew) whose valuations could be cited in legal documents — precisely the role CRI-H100 is designed to fill for GPU collateral. Third, the market took approximately a decade from first issuance to full maturity, suggesting a realistic timeline for GPU-backed structured credit.

2.3.2 Railcar ABS

Trinity Industries' first railcar securitization in 2003 required a credit enhancement (Ambac surety wrap) that would be unnecessary in a mature market. The evolution from credit-enhanced pioneer transactions to standalone, investment-grade-rated ABS demonstrates the path that GPU-backed securities may follow as analytical infrastructure develops (S&P Global, 2008).

2.3.3 Semiconductor Equipment Financing

Jorgenson's research on capital measurement and obsolescence in semiconductor equipment (Jorgenson, 2001) documents how technology-specific depreciation differs from standard economic depreciation. Semiconductor fabrication equipment faces the same fundamental challenge as GPUs: value depends not only on physical condition but on technological relevance relative to newer generations. Jorgenson demonstrates that technology-specific depreciation is empirically measurable but requires asset-class-specific methodology — general-purpose depreciation schedules systematically misstate the actual economic decline.

2.4 Benchmark Governance: Lessons from LIBOR and SOFR

The transition from LIBOR to SOFR provides the most directly applicable governance precedent. LIBOR was a proprietary benchmark administered by a commercial entity (the British Bankers' Association, later ICE Benchmark Administration) with methodology that could be influenced by panel banks with trading positions. The Wheatley Review (2012) and subsequent regulatory action demonstrated that benchmarks embedded in financial contracts require: (a) transparent methodology, (b) independence from commercial interests, (c) formal governance and change control, and (d) independent reproducibility.

SOFR was designed to meet these requirements. It is transaction-based (calculated from actual overnight Treasury repo transactions), transparent (methodology published by the Federal Reserve Bank of New York), and independently verifiable.

CRI-H100 is designed to the same governance standard. The methodology is publicly available, versioned, and subject to formal change control. The data source requires no authentication. Any party can independently reproduce any published value. These are not merely best practices — they are prerequisites for credit document citation in a post-LIBOR regulatory environment.

The CCIR Governance Framework v1.0 is designed with reference to the IOSCO Principles for Financial Benchmarks (July 2013), the international standard for benchmark governance that emerged from the LIBOR reform process. A principle-by-principle alignment summary is included in the Governance Framework (Appendix C). CRI-H100's automated, public-API-based data collection eliminates the category of submitter conflicts that IOSCO Principle 14 addresses — a structural advantage over benchmarks that rely on voluntary submissions from market participants with trading positions.

3. The GPU-Backed Lending Market: Current State

3.1 Market Size and Structure

The GPU-backed lending market has grown rapidly from approximately $2 billion in 2023 to over $20 billion in outstanding obligations by early 2026. Major transactions include CoreWeave ($10.45 billion in GPU-backed debt across multiple facilities), Lambda Labs ($500 million, structured as the first GPU-backed ABS), FluidStack/Macquarie ($150 million with Google as anchor customer), and multiple facilities arranged by BlackRock, JPMorgan, and Carlyle.

Goldman Sachs estimates $736 billion in cumulative AI infrastructure investment by end of 2026. Morgan Stanley projects $2.9 trillion cumulatively by 2028. Morgan Stanley further estimates $680 billion in aggregate chip depreciation charges over the next four years — a figure that frames the scale of the collateral risk embedded in this market.

3.2 The Depreciation Assumption Divergence

Table 1 presents the depreciation assumptions currently observed across market participants. The variance is not a statistical artifact — it reflects fundamentally different views about the economic life of GPU hardware.

Table 1: GPU Depreciation Assumptions Across Market Participants

Entity	Category	Assumed Useful Life	Implied Annual Depreciation	Basis
CoreWeave	Neocloud operator	6 years	~17%	GAAP management estimate
Amazon (revised Feb 2025)	Hyperscaler	5 years	~20%	Revised from 6 years
Meta	Hyperscaler	5.5 years	~18%	GAAP management estimate
Google	Hyperscaler	6 years	~17%	GAAP management estimate
Nebius	Neocloud operator	4 years	~25%	GAAP management estimate
Lambda Labs	Neocloud operator	5 years	~20%	GAAP management estimate
Groq	AI chip company	1 year	~100%	CEO public statement
USD.AI / Hydra	Lending protocol	~6 years (17% p.a.)	17%	Hydra Longevity Study
Aravolta (telemetry)	Monitoring platform	Varies 30–45% by workload	Variable	Fleet telemetry data

Under GAAP (ASC 360-10-35-4), depreciation accounting is “a process of allocation, not of valuation” and useful life is an entity-specific management estimate. The FASB’s Master Glossary defines useful life based on how the entity intends to use the asset, which means two companies operating identical hardware can arrive at different and compliant estimates. This creates a structural problem for credit markets: the depreciation assumption embedded in a loan document is the borrower's management estimate, not an independent market observation.

The "Great Hyperscaler Divergence" of early 2025 — Amazon shortening useful life assumptions for a subset of servers while Meta simultaneously extended its estimates — illustrates that management is using depreciation schedules as earnings management levers under identical technological conditions. This undermines the reliability of GAAP depreciation as a collateral valuation input for credit analysis.

3.3 Collateral Valuation Methods in Current Practice

Our review of available deal documentation, practitioner commentary, and industry reporting identifies five distinct approaches to GPU collateral valuation currently in use:

Credit substitution. The largest transactions (CoreWeave, FluidStack) are underwritten against anchor customer credit quality, not GPU collateral value. CoreWeave's S-1 discloses that DDTL interest rate spreads are explicitly tiered by customer creditworthiness: SOFR + 6.00–6.50% for investment-grade customers versus SOFR + 13.00% for non-investment-grade customers (CoreWeave, Inc., Form S-1, 2025). The collateral enables the loan to exist, but the customer credit determines the pricing. This approach is structurally unavailable to mid-market operators without hyperscaler relationships.

Broker resale quotes. USD.AI and similar lending structures use wholesale GPU brokers (HydraHost, Procurri, Blockware) for point-in-time hardware valuations. This provides real transaction-based data but is proprietary, non-standardized, and lacks forward-looking depreciation modeling.

Internal depreciation models. Each lender develops proprietary assumptions. The Hydra Longevity Study cited by USD.AI is effectively proprietary research from a single broker partner, not an independent market standard.

Subscription data services. Sophisticated lenders subscribe to Silicon Data's index (backed by DRW and Jump Trading), which publishes daily H100 rental rates on Bloomberg. The methodology is proprietary and cannot be independently reproduced.

Abstention. A non-trivial portion of institutional lenders have chosen not to participate. Bridge Bank's CEO told PitchBook: "I don't think anyone really knows the economic shelf life."

None of these approaches constitutes a standardized, independently verifiable, credit-document-referenceable analytical framework. The gap is structural, not informational.

4. The Two-Curve Framework

The central analytical contribution of this paper is the formal separation of two distinct curves that the current market conflates under the single label "depreciation."

4.1 Rental Rate Depreciation (Revenue-Side)

Rental rate depreciation measures how the market price of renting a GPU per hour changes over time as newer architectures arrive, supply expands, and software efficiency evolves. This curve describes the borrower's revenue-generating capacity — the primary source of debt service.

Formally, let r(g, t) denote the market rental rate for GPU generation g at time t. The rental rate depreciation curve for generation g is:

$$\delta\_r(g, t) = [r(g, t) - r(g, t_0)] / r(g, t_0) \tag{1}$$

where t_0 is the reference date (typically the date of loan origination or the first observation in the CRI time series). The cumulative rental rate depreciation at any point provides a direct measure of how much revenue-generating capacity the collateral has lost.

4.2 Hardware Resale Depreciation (Loss-Given-Default)

Hardware resale depreciation measures what a physical GPU sells for on the secondary market over time — the liquidation value in a default scenario. Let v(g, t) denote the secondary market resale price for GPU generation g at time t:

$$\delta\_v(g, t) = [v(g, t) - v(g, t_0)] / v(g, t_0) \tag{2}$$

This curve determines loss given default (LGD) in credit analysis. It is influenced by factors partially independent of rental rates: component harvesting value, secondary market liquidity, geographic demand distribution, and bulk disposal discounts in foreclosure scenarios.

4.3 The Spread Between Curves

The relationship between rental rate depreciation and hardware resale depreciation is analytically important and currently unmeasured. We define the rental-resale spread:

$$S(g, t) = \delta\_r(g, t) - \delta\_v(g, t) \tag{3}$$

When $S > 0$, rental rates have declined faster than resale values — the revenue-side is weaker than the liquidation side. When $S < 0$, resale values have declined faster — the borrower can still service debt but the liquidation recovery in a default scenario is deteriorating faster than the revenue signal suggests.

This spread is a novel risk metric for GPU-backed lending. Under the value cascade thesis (theCUBE Research, 2025), older GPUs retain resale value through repurposing for inference workloads even as rental rates for primary use cases decline. If this thesis holds, S should be persistently positive — rental rate depreciation should outpace resale depreciation — which is favorable for lenders (the liquidation floor remains above the revenue-implied value). If the cascade thesis fails — for example, if next-generation hardware eliminates the inference use case for prior generations — S turns negative, and lenders face simultaneous revenue deterioration and collateral impairment.

Table 2: Two-Curve Framework — Credit Functions

Analytical Input	Curve	Credit Function	Current Source	CCIR Contribution
What the GPU earns	Rental rate depreciation δ_r	Debt service coverage (DSCR)	No open-methodology source	CRI-H100 (Stage 1)
What the GPU sells for	Hardware resale depreciation δ_v	Loss given default (LGD)	Broker quotes (proprietary)	Framework integration
Relationship between curves	Rental-resale spread S	Collateral coverage stress testing	Does not exist	CRI-D (Stage 2)

4.4 Why the Sequence Matters

Stage 2 (depreciation curves) is structurally dependent on Stage 1 (rental rate reference) for three reasons:

Measurement requires baseline. You cannot measure how a rental rate declines over time without first having a consistent, reproducible measure of what the rate is at each point. An ad hoc collection of provider quotes does not constitute a time series — it constitutes noise with a trend.

Cross-generational comparison requires standardized methodology. Depreciation curves are constructed by comparing the price trajectories of successive GPU generations (V100, A100, H100, H200) through their respective technology transitions. These comparisons are only valid if each generation's price trajectory is measured using the same methodology, filters, and governance. Applying CRI methodology consistently across generations creates the controlled comparison necessary for depreciation analysis.

Governance requires track record. A depreciation curve derived from 30 days of data is an illustration. One derived from 12–18 months through a hardware transition cycle is an empirical finding. For credit document citation, the distinction is dispositive. The Stage 1 infrastructure must operate and be validated before Stage 2 outputs carry the credibility required for contract reference.

5. Stage 1: CRI-H100 Index Methodology

This section summarizes the CRI-H100 methodology. The full technical specification is published separately as CCIR Methodology v1.1.0.

5.1 Data Source and Rationale

CRI-H100 is calculated from listings on the Vast.ai GPU rental marketplace. Vast.ai was selected on three criteria:

Reproducibility. The Vast.ai API (https://console.vast.ai/api/v0/bundles/) requires no authentication. Any counterparty, attorney, rating agency, or third party can independently query the same endpoint and reproduce CCIR's inputs without any data licensing agreement. This is the foundational requirement for a credit reference rate in a post-LIBOR environment.

Transparency. Listing data includes provider reliability scores, geographic information, hardware specifications, and pricing — sufficient metadata to apply meaningful, verifiable quality filters.

Market representation. Vast.ai is among the largest public GPU rental marketplaces by listing volume. While it does not represent the entire market (see Section 10, Limitations), it provides the most reproducible proxy available for on-demand spot market pricing.

5.2 Index Classification and Scope

CRI-H100 measures a specific and deliberately narrow quantity. To prevent misapplication in credit documents, we state explicitly what the index is and what it is not.

CRI-H100 is: an on-demand marginal supply rate for secondary marketplace GPU compute capacity. It measures the price at which providers on a public marketplace are willing to supply H100 SXM GPU-hours to any buyer without a pre-existing contract. It is structurally analogous to an electricity spot market price or a posted dealer offer rate — it reflects the marginal cost of acquiring capacity in real time on the open market.

CRI-H100 is not: an enterprise contract benchmark, a volume-weighted clearing rate, a hyperscaler blended rate, or a measure of the average price paid across all GPU rental transactions in the economy. Enterprise contracts between neoclouds and hyperscaler customers (e.g., CoreWeave-Microsoft, FluidStack-Google) are negotiated bilaterally at terms that may diverge substantially from secondary marketplace rates. CRI-H100 does not claim to represent those transactions.

This classification has specific implications for credit applications. CRI-H100 is most directly useful as: (a) a stress indicator — when secondary marketplace rates decline, it signals that uncommitted capacity is becoming cheaper, which pressures enterprise contract renewal rates on a lagged basis; (b) a floor reference — in a distressed scenario where a borrower loses an enterprise contract and must re-lease capacity on the open market, CRI-H100 approximates the rate they would face; (c) a collateral repricing signal — sustained CRI-H100 declines indicate that the revenue-generating capacity of uncommitted GPU inventory is deteriorating, which should trigger collateral revaluation.

Lenders referencing CRI-H100 in covenant structures should calibrate trigger levels to account for the spread between enterprise contract rates and secondary marketplace rates. That spread is itself an informative risk metric: a narrowing spread suggests enterprise pricing is converging toward spot, which may indicate weakening demand or increasing competition.

Market structure context. The GPU compute rental market operates across three distinct pricing tiers: (1) bilateral enterprise contracts between hyperscalers and large neoclouds, negotiated privately with terms that reflect customer creditworthiness, volume commitments, and non-price considerations such as priority allocation and custom configuration — this tier is structurally opaque and not amenable to public index construction; (2) reserved and committed capacity offered by mid-market providers (Lambda, CoreWeave spot, RunPod reserved instances), where pricing is semi-transparent through published rate cards and API-accessible listings; and (3) on-demand secondary marketplace capacity (Vast.ai, RunPod community, and similar platforms), where pricing is fully transparent and independently verifiable. CRI-H100 measures tier 3. Using CRI-H100 to underwrite a facility whose cash flows derive primarily from tier 1 enterprise contracts would be analogous to using retail electricity spot prices to underwrite a power purchase agreement-backed renewables securitization — structurally mismatched. CRI-H100 is appropriately applied to borrowers whose revenue depends on tier 2 or tier 3 pricing, and as a stress-case floor for tier 1 borrowers modeling the loss of an anchor contract.

5.3 Quality Filters

The following filters are applied sequentially to raw listings. A listing must satisfy all filters to become a Qualifying Listing:

Table 3: CRI-H100 Quality Filters

Filter	Criterion	Rationale
GPU model	gpu_name = “H100 SXM”	Consistent hardware specification
Availability	rentable = true AND rented = false	Currently available capacity at listed price
Reliability	reliability2 ≥ 0.90	Excludes unreliable providers
Minimum GPU count	num_gpus ≥ 1	Ensures valid listing
Staleness	start_date within prior 7 days	Excludes stale listings
Geography	geolocation ends with “, US”	US-only scope (Section 5.4)

5.4 Geographic Scope

CRI-H100 v1.1 covers United States listings only. This scope was selected because the majority of GPU-backed credit agreements are governed by US law, US and non-US rental rates may reflect different supply-demand dynamics, and a single-geography index is simpler to interpret and less susceptible to cross-jurisdictional variation. Regional expansion (CRI-H100-EU, CRI-H100-APAC) is planned as separate indices under identical methodology.

5.5 Outlier Removal

Outlier removal prevents anomalous listings from distorting the index. The procedure follows a trimmed-mean sigma method:

Step 1. For each Qualifying Listing, compute per-GPU price: Observation = dph_total / num_gpus.

Step 2. Sort all Observations. Remove the top 10% and bottom 10% by value. Compute the arithmetic mean of the remaining Observations (trimmed mean).

Step 3. Compute the standard deviation of the full (untrimmed) Observation series.

Step 4. Exclude any Observation where |Observation − trimmed_mean| > 2.5 × standard_deviation.

The 2.5σ threshold was selected to balance outlier removal against preservation of legitimate price dispersion. A tighter threshold risks excluding valid market variation; a looser threshold permits manipulative or erroneous listings to influence the index.

5.6 Index Calculation

CRI-H100 is calculated weekly as the median of pooled Observations from a trailing 7-day Calculation Window:

$$\text{CRI-H100}\_t = \text{median}(\{O\_i : i \in W\_t\}) \quad (4)$$

where W_t is the set of all cleaned Observations from days meeting the minimum daily observation threshold (10 Qualifying Listings) in the 7-day window ending on the Wednesday preceding each Thursday publication date.

The median was selected over alternative central tendency measures for specific reasons: (a) manipulation resistance — the median is robust to extreme values and substantially more resistant to manipulation by a small number of anomalous listings than the mean; (b) no volume data dependence — the median requires no weighting by volume, which has uncertain reliability in the current dataset; (c) credit appropriateness — for loan covenant purposes, a robust central tendency is more appropriate than a volume-weighted measure susceptible to large-listing influence.

5.7 Publication and Governance

CRI-H100 is published weekly on Thursdays. Each publication includes the index value (4 decimal places), observation count, valid day count, dispersion metrics (min, max, mean, stdev), a Low Confidence flag if observation thresholds are not met, and the methodology version under which the value was calculated.

The published series is append-only. Once a value is published, it is not revised except under the Restatement provisions of the CCIR Governance Framework v1.0. Each daily data collection includes a SHA-256 hash of the raw data file committed to the public GitHub repository, creating a tamper-evident audit trail.

6. Stage 2: Market-Derived Depreciation Curves

6.1 Construction Methodology

Stage 2 derives empirical depreciation curves by observing how CRI values for successive GPU generations evolve through their respective technology transition cycles. The construction requires:

(a) CRI indices for at least two GPU generations measured under identical methodology (e.g., CRI-H100 and CRI-A100);

(b) Observation of at least one complete technology transition (new architecture launch through rental rate stabilization at the new equilibrium);

(c) Sufficient time series length for statistical significance (minimum 52 weekly observations, i.e., one year of data).

6.2 Historical Analogues and Preliminary Observations

Two prior GPU generation transitions provide empirical precedent. While CCIR's automated data collection was not operational during these transitions, we reconstruct approximate rental rate trajectories from available market data:

V100 → A100 (2020–2022)

The A100 launched in 2020. V100 rental rates declined from approximately $2.48/GPU-hour (Google Cloud on-demand pricing, sustained through 2025 at hyperscaler list prices) to approximately $0.55/GPU-hour on secondary marketplace platforms (Lambda spot pricing, 2025 observation). This represents approximately 78% cumulative decline over approximately 5 years, with the decline concentrated in the first 2–3 years after A100 availability. Critically, the V100 retained positive rental value throughout — consistent with the value cascade hypothesis.

A100 → H100 (2022–2024)

The H100 launched during unprecedented AI demand growth. A100 rental rates declined from approximately $4.10/GPU-hour (AWS p4d.24xlarge launch pricing) to approximately $1.29/GPU-hour (Lambda spot, 2025). This represents approximately 69% cumulative decline over approximately 3 years. CoreWeave reported its A100 fleet "fully booked" at compressed rates, and Azure retirement data shows A100-class infrastructure remaining in production use.

H100 → Next Generation (2025–present)

H100 rental rates have declined from approximately $8–10/GPU-hour at peak (2024) to approximately $2–3/GPU-hour (multiple marketplace observations, early 2026). This represents approximately 60–70% decline in approximately 18 months — faster than either prior transition, likely driven by both supply expansion and the arrival of H200/Blackwell architectures.

6.3 Depreciation Scenario Framework

Stage 2 models three forward scenarios for H100 rental rate depreciation, designed as stress-testing inputs rather than forecasts:

Table 4: Forward Depreciation Scenarios

Scenario	Mechanism	Annual Rental Rate Decline	Implied Useful Rental Life	Historical Basis
A: Gradual cascade	H100 retains inference value as newer architectures absorb training	15–20%	5–6 years	V100 precedent; hyperscaler cascade model
B: Accelerated compression	Next-gen architectures compress rates 30–40% per cycle; limited inference tail	30–40%	3–4 years	A100 precedent under supply expansion
C: Rapid obsolescence	Software efficiency gains + next-gen collapse rental demand	50–70%	1.5–2 years	H100 observed pace; Groq thesis; DeepSeek-class disruption

The scenarios are not mutually exclusive across time. A given GPU generation may follow Scenario A for the first 18 months (high demand, gradual compression), transition to Scenario B as the next generation achieves scale (accelerated compression), and approach Scenario C if a subsequent architecture renders the inference use case uneconomic.

6.4 CRI-D: The Depreciation Index

When sufficient cross-generational data exists (estimated 12–18 months from first CRI-H100 publication), CCIR intends to publish CRI-D — a formal depreciation index measuring observed rental rate decline by GPU generation. CRI-D will use the same open-methodology governance as CRI-H100 and will be designed for direct integration with hardware resale data from broker sources, providing lenders with the complete analytical picture for collateral coverage modeling.

7. Statistical Validation Framework

This section describes the statistical tests to be applied as the CRI-H100 time series develops. We present the framework prospectively; results will be published as data accumulates.

7.1 Representativeness Testing

Coverage ratio. For each collection day, we compute the ratio of Qualifying Listings to total H100 SXM listings on Vast.ai (before quality filters). A sustained coverage ratio below 30% would indicate that quality filters are excluding a disproportionate share of the market, potentially biasing the index.

Observation count stability. We track daily Observation counts over time. Secular decline in Observation counts may signal market migration away from Vast.ai, requiring data source expansion. The Low Confidence flag triggers when daily counts fall below 10 or pooled weekly counts fall below 4.

Price comparison with non-Vast.ai sources. Where available, we compare CRI-H100 values to reported rental rates on other platforms (Lambda, RunPod, CoreWeave spot pricing) to assess whether Vast.ai-derived rates are systematically biased. This comparison is descriptive, not methodological — CRI-H100 is defined as a Vast.ai marketplace rate, and systematic divergence from other platforms is a market structure finding, not a methodology failure.

7.2 Distributional Properties

Normality testing. We apply Shapiro-Wilk and Jarque-Bera tests to the weekly Observation distributions to characterize the price distribution. GPU rental rates are not expected to be normally distributed — we anticipate right-skewness (a tail of high-priced listings) and possible bimodality (reflecting different provider tiers). The median-based calculation is robust to non-normality, but documenting the distribution informs users who may apply parametric methods to the index series.

Stationarity. Augmented Dickey-Fuller (ADF) and KPSS tests will be applied to the CRI-H100 time series to characterize its stochastic properties. We expect the series to exhibit non-stationarity in levels (trending downward due to secular depreciation) but potential stationarity in first differences. This characterization matters for derivative pricing and stress-testing applications.

Volatility clustering. We test for ARCH effects in the CRI-H100 return series using the Engle (1982) LM test. If volatility clusters around specific events (new GPU architecture announcements, major hyperscaler capacity expansions), this informs the design of trigger mechanisms in credit documents.

7.3 Manipulation Resistance

A reference rate embedded in credit documents faces manipulation incentives that a purely informational index does not. We assess manipulation resistance through four approaches, the last of which — quantitative simulation — is designed to produce empirical evidence of the index’s robustness under adversarial conditions.

Market depth relative to index influence. We quantify the minimum capital required to move the median by a given percentage. Because the index uses the median of all Qualifying Listings (not a volume-weighted average), manipulation requires controlling the majority of listings — a substantially higher threshold than manipulating a mean or a volume-weighted index.

Benford's Law analysis. We apply first-digit frequency analysis to Observation series to screen for artificial price clustering, which could indicate coordinated listing behavior. Deviations from the expected Benford's distribution trigger investigation but are not dispositive (GPU rental prices have structural reasons for non-Benford distributions, including pricing conventions around round dollar amounts).

Day-over-day jump analysis. We monitor the frequency and magnitude of day-over-day jumps in the daily median. Jumps exceeding 3σ from the daily return mean are flagged for investigation. The Manifest Error Override provision in the CCIR Methodology allows the Administrator to exclude clearly erroneous observations that survive automated outlier removal, subject to documentation and public disclosure.

7.3.1 Adversarial Injection Simulation

To quantify manipulation resistance empirically, we conduct the following simulation using actual daily collection snapshots once the 30-day burn-in period is complete:

Protocol. For each daily snapshot containing N Qualifying Listings, we inject k adversarial listings at a target price p_adv and recompute the median. We vary k from 1 to N/2, and we set p_adv at both the upward attack level (150% of the true median) and the downward attack level (50% of the true median). The simulation produces the minimum k required to shift the published median by 5%, 10%, and 25%.

Cost estimation. Each injected listing corresponds to a real economic cost: the adversary must create and maintain a Vast.ai provider instance with a qualifying reliability score (≥ 0.90), which requires actual GPU hardware serving real rentals over time. The simulation translates the required k into an estimated minimum capital outlay based on current H100 hardware costs and the reliability-building period required to pass quality filters.

Expected results. For a daily snapshot with N = 15 Qualifying Listings, the median requires displacement of at least 8 listings (more than half). Under the current quality filter regime, creating 8 new qualifying providers with reliability ≥ 0.90 requires sustained operation over multiple weeks with genuine rental activity — a meaningful economic barrier. For a 7-day pooled weekly calculation with observations drawn from multiple daily snapshots, the manipulation threshold increases proportionally.

Table 5: Illustrative Manipulation Cost Thresholds (to be populated with empirical data)

Daily Observations (N)	Listings to Shift Median 5%	Listings to Shift Median 10%	Listings to Shift Median 25%	Est. Minimum Capital Outlay
8	TBD	TBD	TBD	TBD
12	TBD	TBD	TBD	TBD
15	TBD	TBD	TBD	TBD

20	TBD	TBD	TBD	TBD
30+	TBD	TBD	TBD	TBD

Results from this simulation will be published as Appendix A to the first CRI-H100 publication. The simulation code will be included in the open-source repository so that any counterparty can replicate the analysis on any daily snapshot.

Comparison with alternative estimators. We additionally compare manipulation resistance across four estimators — median, trimmed mean, volume-weighted mean, and Hodges-Lehmann estimator — to confirm that the median is optimal for the observed data characteristics. If an alternative estimator demonstrates superior robustness without sacrificing interpretability, CCIR will consider a methodology revision through the formal change control process.

7.4 Reproducibility Verification

The index is designed for independent reproducibility. The verification protocol is:

(a) Third party queries the Vast.ai API using the parameters specified in CCIR Methodology v1.1.0.

(b) Third party applies the quality filters and outlier removal specified in the methodology.

(c) Third party computes the weekly median.

(d) Third party compares the reproduced value to the published value.

A verification script (pipeline/verify.py) is published in the open-source repository. CCIR commits to making the full raw data and metadata available for each collection day, along with SHA-256 hashes recorded at the time of collection for tamper detection.

Reproducibility may be imperfect due to API timing — a third party querying the API at a different time than CCIR's daily collection may observe a different snapshot. This limitation is disclosed in the methodology and is inherent to any real-time marketplace data source. The SHA-256 hash of CCIR's actual snapshot provides a verifiable record of the exact data used in each calculation.

7.5 Confidence Analysis Under Thin Observation Regimes

CRI-H100 operates in a thin market. Early-period daily observation counts of 8–20 Qualifying Listings are substantially below the data depth available to established commodity or financial benchmarks. Rather than treating this as an undisclosed weakness, we quantify its impact on index precision.

7.5.1 Bootstrap Confidence Intervals

For each weekly publication, we compute a nonparametric bootstrap confidence interval around the published median. The procedure is:

(a) From the pooled weekly Observation set of size N, draw B = 10,000 bootstrap resamples of size N with replacement.

(b) Compute the median of each resample.

(c) Report the 2.5th and 97.5th percentiles of the bootstrap median distribution as the 95% confidence interval.

This confidence interval is published alongside each weekly CRI-H100 value, giving users an explicit measure of statistical precision. In thin observation regimes, the interval will be wide — and that width is itself informative: it tells a lender how precisely the market rate is currently estimable.

7.5.2 Precision Sensitivity by Observation Count

We model expected confidence interval width as a function of observation count to provide forward guidance on index precision:

Table 6: Expected 95% Confidence Interval Width by Observation Count

Weekly Observations (N)	Expected 95% CI Width (% of median)	Interpretation	CCIR Classification
< 20	Wide (≈15–25%+)	Directionally informative; not suitable for tight covenant triggers	Low Confidence
20–40	Moderate (≈8–15%)	Usable for broad covenant structures with appropriate buffers	Standard
40–80	Narrowing (≈4–8%)	Suitable for most credit applications	Standard
80+	Tight (≈< 4%)	Suitable for mechanical triggers and structured product waterfall reference	High Confidence

These estimates are illustrative pending empirical calibration from the burn-in data. Actual confidence interval widths depend on the dispersion of the underlying price distribution, not only on N. The width estimates will be updated with empirical data and published as a standing table in each quarterly CRI methodology review.

The practical implication is that covenant structures referencing CRI-H100 should be designed with awareness of the current confidence regime. A cash sweep trigger set at a 5% decline is appropriate when the confidence interval is tight; the same trigger is premature when the confidence interval is 20%. CCIR publishes the confidence interval explicitly so that lenders and structurers can calibrate accordingly.

7.5.3 Minimum Viable Observation Count

We define a minimum viable observation count (MVOC) as the N below which the 95% confidence interval exceeds 25% of the published median. Below the MVOC, the index value is published with a Low Confidence flag indicating that the statistical precision is insufficient for mechanical covenant reference. The MVOC will be empirically determined during the burn-in period and published in the first quarterly methodology review.

The MVOC is not a hard publication cutoff. CRI-H100 publishes in all weeks where data exists, flagging precision limitations rather than suppressing data. Suppression would create information gaps that are worse for credit markets than wide confidence intervals — a lender who knows the rate is approximately \$2.00 ± \$0.40 is better positioned than a lender who has no data point at all.

8. Credit Market Applications

8.1 Covenant Design

CRI-H100 enables mechanical, non-discretionary covenant structures anchored to an independent reference rate. Below we present four covenant templates with the economic rationale for each.

8.1.1 Minimum Rental Rate Floor

> *The Borrower shall maintain a weighted average rental rate on the Pledged GPU Assets of not less than [●]% of the Rental Rate Reference (CRI-H100) for the trailing 4-week period.*

Economic function: Detects collateral underperformance relative to market. If the borrower's GPUs rent below market — due to poor contracts, low utilization, or hardware degradation — the lender receives an early warning before debt service coverage deteriorates.

8.1.2 Loan-to-Value Maintenance

> *The Collateral Value shall be calculated as [●] × CRI-H100 × aggregate GPU-hours of available capacity per annum. The Borrower shall maintain LTV not greater than [●]%.*

Economic function: Creates a dynamic, market-linked LTV test. As CRI-H100 declines, calculated collateral value declines, triggering cure obligations before the lender is substantially underwater.

8.1.3 Cash Sweep Trigger

> *If CRI-H100 falls below $[●]/GPU-hour for [2] consecutive Publication Dates, excess cash flow above debt service shall be swept to a reserve account.*

Economic function: Automatic cash conservation without requiring a default declaration. Analogous to excess spread triggers in auto loan ABS, but linked to an external market reference rather than pool-level delinquency.

8.1.4 Appraisal Event Trigger

> *If CRI-H100 declines by more than [25]% from the Closing Date level over any trailing [90]-day period, the Borrower shall deliver an independent appraisal within [30] Business Days.*

Economic function: Rapid market rate compression triggers mandatory reappraisal before material LTV deterioration. Structurally analogous to automatic appraisal triggers in CMBS.

8.2 ABS Structuring Applications

For GPU-backed securitizations, the two-stage framework provides inputs at each stage of the rating and structuring process:

Collateral valuation. CRI-H100 provides a standardized, auditable methodology for marking GPU collateral to market — necessary for computing overcollateralization ratios and pool-level coverage tests.

Depreciation stress scenarios. The three-scenario framework (Section 6.3) provides structured inputs for rating agency sensitivity analysis: base case (Scenario A), downside (Scenario B), and severe stress (Scenario C).

Performance monitoring. Published weekly, CRI-H100 gives servicers, trustees, and investors an ongoing collateral performance reference after closing — analogous to appraisal updates in CMBS or servicer reports in auto ABS.

Trigger design. Early amortization, cash sweep, and performance event triggers can be mechanically linked to CRI-H100 levels, reducing servicer discretion and corresponding moral hazard.

8.3 Rating Agency Interface

The framework is designed to support rating agency criteria development for GPU-backed structured products. CCIR anticipates that rating agencies will require: (a) an independent, auditable data source for collateral valuation; (b) depreciation scenarios for sensitivity analysis; (c) methodology governance meeting post-LIBOR documentation standards; and (d) a commitment to independent audit prior to closing of any rated transaction referencing CRI-H100.

The CCIR Governance Framework v1.0 addresses each of these requirements. The Rating Agency Disclosure Statement in Section 15 of that document provides a structured summary of index characteristics, material risk disclosures, and known limitations formatted for rating agency evaluation.

9. Intra-Model Performance Variance and Collateral Heterogeneity

9.1 Manufacturing Yield Variance in GPU Performance

Research published at the GPGPU ’26 conference (Silicon Data, March 2026) documents intra-model GPU performance variation of up to 38% within the H100 SXM model across cloud providers. The variation is attributed to semiconductor manufacturing yield variance — known informally in the hardware community as the “silicon lottery” — as well as provider configuration differences, thermal management, and workload tuning. The study encompassed 3,500+ GPUs across multiple cloud providers with 6,800+ benchmark runs.

This finding has direct implications for any GPU rental rate index, including CRI-H100: two listings at identical $/GPU-hour may represent economically different products. A buyer paying $2.00/hour for an H100 at the 95th percentile of the performance distribution receives materially more compute per dollar than a buyer paying $2.00/hour for an H100 at the 5th percentile.

9.2 Impact on CRI-H100

CRI-H100 v1.1 does not adjust for performance variance. This is a deliberate design decision:

(a) Performance grading requires benchmark data that CCIR does not currently produce and that cannot be obtained from the Vast.ai API. Incorporating performance adjustment would require either licensing proprietary benchmark data (creating a dependency inconsistent with the open-methodology commitment) or running independent benchmarks (operationally beyond current scope).

(b) A price index that explicitly discloses what it measures and what it does not is more analytically useful than a performance-adjusted index with opaque normalization methodology.

(c) The performance variance finding is itself subject to further research and validation. Embedding a performance adjustment based on a single study would create methodology fragility.

9.3 Implications for Credit Analysis

Intra-model performance variance has specific implications for GPU collateral valuation that lenders should consider independently of whether a price index adjusts for performance:

Collateral heterogeneity. A fleet of 1,000 H100s is not 1,000 identical assets. The fleet’s aggregate performance — and therefore its aggregate revenue-generating capacity — depends on the performance distribution of the individual units. Aravolta’s telemetry data indicates that fleet-level depreciation curves vary 30–45% depending on workload profiles, even for identical hardware models.

Performance certification. The presence of significant intra-model performance variance suggests that lenders may benefit from requiring standardized performance certification at origination, analogous to environmental site assessments in real estate finance or condition surveys in aviation lending. A lender should know the performance profile of the collateral at origination and should have the ability to re-test at covenant dates. Without such certification, the lender cannot distinguish between a fleet of high-performing units and a fleet of low-performing units — even if both fleets contain the same number of nominally identical GPUs.

9.4 Roadmap: CRI-H100-PA

CCIR is developing a performance-adjusted variant (CRI-H100-PA) as a separate index. Publication is contingent on establishment of a reproducible performance grading methodology. CRI-H100-PA will complement, not replace, the price index — the two indices answer different questions and both are useful for credit analysis.

10. Limitations and Areas for Further Research

10.1 Single Data Source

CRI-H100 v1.1 is based solely on Vast.ai marketplace data. Enterprise contracts, hyperscaler pricing, private capacity agreements, and other marketplace rates are not reflected. Vast.ai represents a reproducible proxy for on-demand spot pricing but does not capture the full market. Observation counts on the US H100 SXM segment of Vast.ai are currently thin (8–20 Qualifying Listings per day in early 2026). CCIR publishes observation counts and applies the Low Confidence flag when thresholds are not met. Data source expansion is a roadmap priority, subject to the requirement that any additional source must preserve independent reproducibility.

10.2 Listed Prices vs. Executed Prices

CRI-H100 measures listed rental rates, not executed transaction prices. A listing appearing at $1.80/GPU-hour may or may not be rented at that price. The index reflects willingness to supply, not confirmed transactions. This limitation is structural to any open-data index at this stage — executed transaction data requires proprietary data relationships incompatible with the open-methodology requirement. Ornn's OCPI, which uses executable offer data, provides a partial solution but at the cost of reproducibility.

10.3 Historical Data Limitations

The V100 and A100 depreciation analogues presented in Section 6.2 are reconstructed from publicly available pricing data (hyperscaler list prices, marketplace spot rates), not from a contemporaneous CRI-equivalent index. These reconstructions involve estimation and interpolation. CRI-H100 will produce the first contemporaneous, methodologically consistent rental rate record for any GPU generation — but only prospectively from the collection start date.

10.4 Performance Variance

As discussed in Section 9, CRI-H100 does not adjust for intra-model performance variance. Users interpreting CRI-H100 as a measure of compute value per dollar (rather than price per GPU-hour) should apply appropriate adjustments for the performance distribution of the specific collateral under analysis.

10.5 Technology Discontinuity Risk

The depreciation scenarios in Section 6.3 model gradual-to-rapid rental rate decline. They do not model discontinuous technology events that could render GPU compute fundamentally uncompetitive for current use cases: quantum computing achieving practical advantage for AI workloads, novel non-GPU architectures (e.g., photonic computing, neuromorphic chips) displacing GPU demand, or algorithmic advances that dramatically reduce compute requirements. These events are unlikely on the 3–5 year horizon relevant to most GPU-backed loans but represent tail risks that the scenario framework cannot capture.

10.6 Areas for Further Research

Several extensions warrant further investigation:

Multi-source index construction. As additional marketplaces develop public APIs or enter data-sharing agreements, CRI methodology should incorporate multiple sources while preserving reproducibility. The weighting methodology for multi-source indices requires careful design to prevent source concentration risk.

Performance-adjusted indices. As discussed in Section 9.4, development of CRI-H100-PA requires a standardized, reproducible performance grading methodology. Collaboration with Silicon Data or independent benchmark development are both under evaluation.

Cross-regional spread dynamics. Publication of CRI-H100-EU and CRI-H100-APAC indices would enable formal analysis of geographic price structure — the compute analogue to locational marginal pricing. Preliminary evidence suggests persistent regional price divergence, but formal measurement awaits regional index publication.

Rental-resale spread analysis. Formal measurement of the S(g,t) spread defined in Section 4.3 requires pairing CRI rental rate data with time-matched hardware resale data from broker sources. This integration is analytically valuable and operationally feasible but requires a data-sharing relationship with at least one hardware broker.

Forward curve construction. Development of term structure models for GPU rental rates — analogous to the electricity forward curve — would enable derivative pricing, hedging instrument design, and forward-looking collateral valuation. This requires a liquid spot market as its foundation, which is the primary contribution of Stage 1.

Reserved-capacity reference rate. CRI-H100 measures tier 3 (on-demand secondary marketplace) pricing as described in Section 5.2. A complementary index measuring tier 2 (reserved and committed capacity from mid-market providers) would serve borrowers whose revenue profiles are tied to reserved-instance pricing rather than spot. Construction of such an index requires API access to published reserved-instance rate cards from multiple providers, or data-sharing agreements with mid-market platforms. The methodology would follow the same open, reproducible principles as CRI-H100 but would capture a different segment of the pricing distribution — one more relevant to the mid-market borrowers who represent the most likely early adopters of CRI-referenced credit structures.

11. Conclusion

The GPU-backed credit market is growing rapidly into a gap in its own analytical infrastructure. Depreciation assumptions vary by a factor of six. No independently verifiable rental rate reference exists. No market-derived depreciation curves have been published. The consequence is an asset class where lenders either substitute credit for collateral, charge substantial premiums for unquantified uncertainty, or abstain entirely.

This paper proposes a structured response: a two-stage framework in which a transparent rental rate reference is established first, and market-based depreciation curves are constructed from the resulting time series second. The framework draws on established precedents from commodity financialization, electricity market design, and equipment-backed securitization to argue that the approach is both structurally sound and practically implementable.

The contribution is deliberately narrow. CRI-H100 measures one thing — listed H100 SXM rental rates on one marketplace in one geography — and measures it with full transparency and independent reproducibility. The depreciation framework integrates this measurement with hardware resale data from broker sources, producing the complete analytical picture lenders need. Intra-model performance variance is disclosed honestly and addressed on the roadmap. The governance structure is designed with reference to the IOSCO Principles for Financial Benchmarks and meets post-LIBOR standards for credit document citation.

Every major asset-backed market developed this kind of analytical infrastructure. Aircraft finance has AVITAS appraisals and EETC structural conventions. Real estate has CMBS appraisal methodologies and S&P/Moody's criteria. Auto ABS has standardized pool-level performance data. GPU-backed lending currently has none of these.

The question is not whether this infrastructure will be built. It is whether it will be built before or after the first credit stress event in the GPU-backed lending market. The refinancing wall facing 2023–2025 vintage GPU loans, the UBS tail risk scenario for AI-driven private credit disruption, and the continued rapid decline in H100 rental rates all suggest that the window for proactive infrastructure development is measured in quarters, not years.

CCIR's contribution is the first stage of that infrastructure: an open-methodology, independently verifiable rental rate reference designed specifically for credit market use, published with the governance and documentation standards that structured finance requires. The second stage follows as the data accumulates. Both are needed. The sequence matters.

Companion Documents

CCIR Methodology v1.1.0 | CCIR Governance Framework v1.0

https://github.com/ccir-index/ccir-methodology

References

Borenstein, S., Bushnell, J., & Wolak, F. (2002). Measuring market inefficiencies in California's restructured wholesale electricity market. American Economic Review, 92(5), 1376–1405.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987–1007.

Bloomberg News (2026). CoreWeave seeks $8.5 billion loan from banks backed by Meta deal. Bloomberg, February 2026.

CoreWeave, Inc. (2025). Form S-1 Registration Statement. SEC File No. 333-283527, filed March 2025.

CoreWeave, Inc. (2025). Form 8-K: Fifth Amendment to DDTL 2.0 Credit Agreement. Filed September 29, 2025.

Gorton, G. (2010). Slapped by the invisible hand: The panic of 2007. Oxford University Press.

Gorton, G., & Rouwenhorst, K.G. (2006). Facts and fantasies about commodity futures. Financial Analysts Journal, 62(2), 47–68.

Hogan, W.W. (1992). Contract networks for electric power transmission. Journal of Regulatory Economics, 4(3), 211–242.

IOSCO (2013). Principles for Financial Benchmarks: Final report. International Organization of Securities Commissions, FR07/13, July 2013.

Jorgenson, D.W. (2001). Information technology and the U.S. economy. American Economic Review, 91(1), 1–32.

Joskow, P.L., & Tirole, J. (2000). Transmission rights and market power on electric power networks. RAND Journal of Economics, 31(3), 450–487.

Littlejohns, A., & McGairl, S. (2006). Aircraft securitisation and leasing. Euromoney Books.

Mish, M. (2026). AI disruption and private credit: A tail risk scenario. UBS Credit Strategy Research.

S&P Global (2008). Criteria for rating railcar securitizations. S&P Global Ratings.

Silicon Data (2026). Intra-model GPU performance variance: Implications for compute pricing and collateral valuation. Proceedings of GPGPU '26.

Tang, K., & Xiong, W. (2012). Index investment and the financialization of commodities. Financial Analysts Journal, 68(6), 54–74.

The Economist (2026). The race to create a financial market for computing power. The Economist, February 2026.

theCUBE Research (2025). Resetting GPU depreciation: Why AI factories bend, but don't break, useful life assumptions. SiliconANGLE/theCUBE Research, Breaking Analysis No. 298.

Wheatley, M. (2012). The Wheatley Review of LIBOR: Final report. HM Treasury, United Kingdom.