

GPU RESIDUAL VALUE 2026 ONE-PAGER

The condensed one-page version of the full 12-page report found [here](#).

American Compute

WHY RESIDUAL VALUE MATTERS

In AI infrastructure, the GPU residual value, or resale price, impacts every part of the economics, across leases, loans, equity returns, and tax treatment, collectively influencing the flow of trillions of dollars.

Most lenders agree that over a three-year term a GPU server should fetch more than 5% of original cost; the debate is whether the right number is 10% or 60%.

It is our belief that GPU residuals are underwritable. We have transaction data covering over 600,000 secondary-market units and run rigorous scenario analysis with actuarial firms and leading reinsurers.

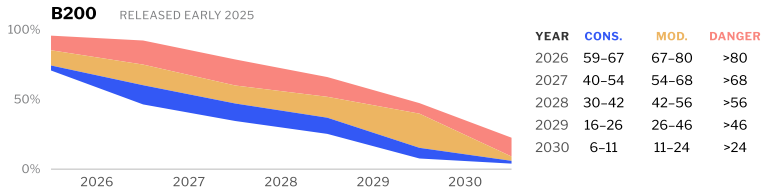
We believe a GPU's useful life can reach **eight years**. We also believe investors and lenders can book meaningful residuals, even **north of 10%** of equipment cost over five years.

THE RESIDUAL BANDS

We publish GPU residuals as three bands rather than a single curve.

Conservative	The safe and reliable scenario.
Moderate	Aggressive but potentially defensible.
Dangerous	Very unlikely, bookable only when everything goes right.

Exhibit 1: Our outlook on GPU residuals as of June 2026



Other GPUs are in the full report. Table values are mid-year sampled band percentages of original MSRP.

WHAT HOLDS VALUE

OEM brands Dell-built servers resell about 18% above less-known OEMs.	Whole vs. parted-out A whole tested, configured cluster sells better than nodes sold individually.
Vanilla vs. exotic Standard configurations clear faster and higher than bespoke ones.	Geography GPUs deployed in the U.S. resell better than identical units abroad.
Documentation Service records and original configuration documents help.	Timing Give the resale process 30 to 60 days; a rushed sale fetches a lower residual.

MORE THAN JUST GPUS

The GPU shortage is over.	Exhibit 4: GPUs are ~75% of the bill of materials														
What is scarce in mid-2026 is everything <i>around</i> the GPU: grid interconnection, datacenter shells, memory, and storage.	<table><thead><tr><th>COMPONENT</th><th>COST SHARE</th></tr></thead><tbody><tr><td>GPU</td><td>76.8%</td></tr><tr><td>SSD</td><td>7.9%</td></tr><tr><td>NIC</td><td>5.7%</td></tr><tr><td>IB / Switches</td><td>4.8%</td></tr><tr><td>CPU</td><td>3.8%</td></tr><tr><td>RAM</td><td>1.0%</td></tr></tbody></table>	COMPONENT	COST SHARE	GPU	76.8%	SSD	7.9%	NIC	5.7%	IB / Switches	4.8%	CPU	3.8%	RAM	1.0%
COMPONENT	COST SHARE														
GPU	76.8%														
SSD	7.9%														
NIC	5.7%														
IB / Switches	4.8%														
CPU	3.8%														
RAM	1.0%														
This has led to a temporary increase in residuals of older generation equipment: it's actually easier to buy new B200s than used H100s.	Source: American Compute reference BOM weights														

MODELING PITFALLS

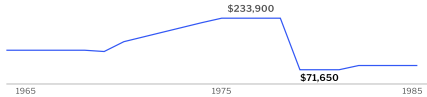
Indices and rental rates Only an arm's-length secondary transaction prices the asset itself.	DCF as the basis Rental-rate volatility bleeds straight into the analysis.
Overthinking physical wear Buyers don't care about a single GPU's thermal history.	Curves instead of distributions A single curve hides what underwriting cares about: dispersion.
Consumer vs. enterprise Consumer comps have nothing to do with how enterprise systems clear.	Depreciation schedules as residuals Tax and reporting conventions fail to predict residuals.
Asset age instead of technology age A brand-new H200 is already more than two years into its technology life.	Overlooking delays A Blackwell project signed today might only be racked in six months.

THE OVERBOOKING TRAP

Overbooking, or booking a residual higher than it is worth, is common during speculative periods. A high residual prices the lease lower, lets the loan advance more, and lets IRR projections seem higher.

In 1979, IBM cut its own pricing on its newest mainframes due to mounting competition and an "overly healthy" secondary market. The used market repriced immediately. Losses ran into the billions in today's dollars.

Exhibit 2: IBM cut entry prices 69%



Sources: Datapro reports; IBM announcement archives. Public entry-configuration purchase prices, nominal USD.

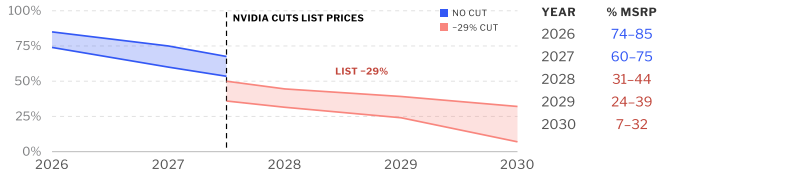
The biggest risk might come from NVIDIA itself. In 1979, IBM held roughly 70% of the computer market. In 2026 NVIDIA ships roughly 90% of datacenter GPUs, at above 70% gross margin. NVIDIA could decide to do the same if pressured.

SCENARIO MODELING: FIVE ADVERSE EVENTS

Two of them can break the moderate band, and potentially the conservative band.

HIGH	NVIDIA margin compression	Mounting pressure sets up NVIDIA to take the same steps that IBM did in 1979.
HIGH	AI demand pop	A 2026 enterprise survey from Cast AI put GPU utilization at 5%.
MEDIUM	Bullwhip over-ordering	Double-orders surface as supply at exactly the moment demand normalizes.
MEDIUM	Hyperscaler refresh	A single refresh event could release billions of dollars of used hardware at once.
LOW	Technology obsolescence	Prior-generation GPUs keep producing economic value through a use-case ladder.

Exhibit 3: If NVIDIA cut gross margins from 75% to 65%, residuals could slip 29%



Illustrative scenario, not a forecast. An estimated ~29% residual depression would happen if NVIDIA decided to cut prices to roughly a 10% premium over AMD's MI-series.

THE POWER AND COOLING WALL

Past roughly 40 kilowatts per rack, air cooling stops working and direct-to-chip liquid cooling becomes mandatory.

So even as faster silicon arrives, the power and cooling wall keeps the prior generation deployable for years.

Hoppers are unique: they are one of the last widely adopted air-cooled, lower-power GPUs.

Exhibit 5: Power per GPU keeps climbing

■ AIR-COOLED
■ LIQUID COOLING RECOMMENDED

GPU Model	Power (W)	Cooling Type
H100	700 W	AIR-COOLED
B200	1,000 W	AIR-COOLED
Rubin	2,300 W	LIQUID COOLING RECOMMENDED

Sources: NVIDIA; SemiAnalysis.

CONCLUSION

- GPU residuals will not always follow a curve. And not everyone quoting an overly high residual is being honest with themselves.
- Model in bands, not a curve.**
Pick a defensible band, not a point. Compare your financial performance at the top and bottom of the range.
 - Pick the band that fits your risk appetite.**
Lessors should sit at the most conservative end, lenders can be slightly more aggressive, and equity investors can be the most aggressive of all. The biggest danger is every party using the same band.
 - Add confidence where it matters.**
Beyond the residual value insurance we structure, a few market solutions can backstop or guarantee your underwritten residual.

American Compute structures residual value insurance solutions for GPU servers.

We work with A-rated reinsurers and can structure solutions for transactions up to **\$500M**. Where needed, we also consult and provide appraisals. Contact us at deals@amcompute.com.

Source: American Compute proprietary dataset — 76,775 completed secondary-market transactions, covering 622,098 units across 679 components, collected continuously since 2023. Arm's-length sales only; no asking prices, no rental rates, no appraisal indices. The residual is measured against original MSRP, after outliers are trimmed. This document is provided for informational and educational purposes only and does not constitute investment, insurance, tax, accounting, or legal advice; it should not be relied upon in making any investment, financing, or underwriting decision. © 2026 American Compute, Inc. All rights reserved.