Episode Book

EPISODE #6007

What a $1M AI Inference Contract Actually Contains

PUBLISHED	15 October 2026 at 19:04 UTC
DURATION	26:36
LISTEN	myweirdprompts.com/episode/ai-inference-contract-negotiation-97150/
HOSTS	Corn, Herman, Jim, Hilbert
SCRIPT MODEL	deepseek-flash
VOICES	chatterbox-regular (chatterbox-local)
PIPELINE	V5.5
SOURCES	14 cited by the research (listed at the end)
BOOK GENERATED	19 October 2026 at 17:42 UTC

SYNOPSIS

Enterprise AI inference isn't bought like seats or like cloud compute — it's a hybrid of both, and the gap between the published rate card and the signed contract is where the entire market lives. This episode walks through what actually happens once you cross the threshold: the commit floors that get a provider to pick up the phone, the discount curves that disagree by twenty points between benchmarks, and the take-or-pay clauses that leave a median 42% of committed credits unspent. We look at how Anthropic and OpenAI diverge on overage policy, why Azure provisioned throughput reservations don't actually reserve capacity, what the cloud overlay is really worth, and why twelve-month exit rights may matter more than any discount percentage on the table.

TRANSCRIPT



Corn

Here's the question that started this one. What does a million-dollar AI bill actually look like on paper?



Herman

And the answer is that the published token price you see on the website is the sticker in the window. It's not the deal on the table.



Corn

Daniel wants to know what happens between those two numbers. He's asking where enterprise negotiation actually begins, what a large inference contract contains once you're past the threshold, how discounts and commitments and service levels and overage are structured, how much leverage a big buyer really has when they threaten to move workloads between the labs.



Herman

And the cloud question. Whether buying through Azure or Bedrock changes the math.



Corn

And what happens when consumption blows past the commitment. Because that's where the labs diverge.



Herman

The gap between the rate card and the contract. That's where the whole market lives. Everything Daniel's asking about sits in that gap.



Corn

So the first thing to understand is that enterprise AI inference isn't bought like seats, and it isn't bought like cloud compute either. It's a hybrid.



Herman

It borrows the seat model's commitment structure and the compute model's metering. You commit to a floor, you draw down against it, and you pay for what you actually burned, up to and past the floor.



Corn

But there's a lot of money in this now. How much are we actually talking about?



Herman

More than a hundred firms spent over ten million dollars each with Anthropic in the twelve months to June of this year. More than a thousand spent over a million. And the average Fortune 500 organization spent nine point four million on external AI platform services in 2025, growing over sixty percent annually.



Corn

So this isn't a niche. This is a real procurement discipline that's still being invented.



Herman

That's the tension that makes it interesting. Token prices have fallen forty to eighty percent over twenty-four months. But multi-year commitments lock in rates. So the structure of the deal determines whether you're hedged or trapped.



Corn

Same product, two outcomes, depending entirely on a clause.



Herman

Exactly the thing Daniel's poking at. So where does the negotiation actually start?



Corn

Start with the floor. At what point does a provider even pick up the phone?



Herman

The advisory benchmark from SaasDash puts it at roughly ten to fifteen thousand dollars a month per provider. Below that, you're on the rate card, full stop. There's no meaningful conversation to have.



Corn

That's monthly spend, not annual.



Herman

Monthly spend per provider. But that's the entry point for a conversation, not the entry point for a real discount. The real enterprise tier starts higher. VendorBenchmark puts the minimum for meaningful OpenAI volume discounts at five hundred thousand dollars in annual committed spend. Anthropic's enterprise tier starts at two hundred and fifty thousand.



Corn

So there's a band between fifteen thousand a month and half a million a year where you can talk, but you're not going to get much.



Herman

You'll get something. You won't get the good curve.



Corn

What does the good curve look like? Give me the shape of it.



Herman

VendorBenchmark's OpenAI benchmarks run like this. Two hundred fifty to four hundred ninety-nine thousand in annual commit, nine to twelve percent off. Five hundred to nine hundred ninety-nine thousand, eleven to seventeen. One million to two point four million, fourteen to nineteen. Two point five to four point nine million, eighteen to twenty-three. And five million plus, twenty to twenty-five, with twenty-five described as the market ceiling.



Corn

Twenty-five percent. That's the top of the mountain.



Herman

That's the top of the mountain per that benchmark.



Corn

And yet I keep seeing much bigger numbers quoted.



Herman

That's the thing. SaasDash gives materially higher ranges on the same question. One hundred to five hundred thousand in commit, twenty to thirty percent. Five hundred thousand to a million, thirty to forty. A million plus, forty to fifty.



Corn

So one benchmark says twenty-five percent is the ceiling and the other says you start above that.



Herman

That disagreement is itself the story. Nobody has a gut feel for a fair price per million tokens yet. And that's exactly why the vendor has room to quote high.



Corn

If I'm the buyer and I have two benchmarks that disagree by twenty points, I have no anchor.



Herman

You have no anchor. And the vendor knows it. So the benchmark is the only reference you have, and the only way to make it credible is to run the evaluation that produces your own number.



Corn

What does Anthropic look like on this? Different shape?



Herman

Anthropic discounts on large commits run around fifteen percent off list, per reporting in The Information. They build the discount as a curve from commit bands, model mix, and term length, and the slope flattens at the top. So the tenth million in commit buys you less discount than the first.



Corn

Term length feeds the discount, does it?



Herman

Term length feeds it. Buyers should pair longer terms with price protection and reforecast rights, though — otherwise you've traded optionality for a point or two.



Corn

For how many years?



Herman

Typically one to three. And that's the product structure now. OpenAI's Guaranteed Capacity product is the canonical template here. One to three year commitments, discounts rising by annual commitment, certainty of access to compute by spend level, drawdown across the OpenAI product portfolio, and use across supported cloud providers and model families.



Corn

So you're not committing to a model. You're committing to a portfolio.



Herman

You're committing to a spend level and a set of access guarantees, and you draw against it across everything OpenAI sells.



Corn

Which means the commit is closer to a treasury instrument than a purchase order.



Herman

That's the right way to think about it. And it's why what happens at the end of the commit matters so much.



Corn

The unused credits. What does the data say about how much gets left on the table?



Herman

Redress Compliance measured median unused credits at the end of the commit year at forty-two percent.



Corn

Forty-two percent.



Herman

Forty-two percent of the commitment, unspent, still owed. Take-or-pay is the core risk. You commit five hundred thousand, you use three hundred fifty, you still owe five hundred.



Corn

That's a two-hundred-thousand-dollar donation to the vendor for the privilege of having been optimistic.



Herman

And their finding on how to avoid it is interesting. Buyers who committed at the seventy-fifth percentile of their trailing pilot consumption — so below what the pilot suggested they'd need — avoided year-end waste. The pilot overstates. It always overstates, because everyone runs their best-case workload on it.



Corn

So the advice is commit below your own pilot, which is the opposite of what the sales motion wants.



Herman

The sales motion wants you at the pilot plus growth. And the protections you negotiate are what close that gap. Ramp schedules, so Q1 runs at sixty percent of the annual pace and Q4 at a hundred and twenty. A one-time downward adjustment window, twenty to twenty-five percent. Rollover credits. Rate stability.



Corn

Which of those actually matters most?



Herman

The downward adjustment window. It's the only one that lets you admit the forecast was wrong in writing, once, without renegotiating the whole agreement.



Corn

Hm.



Herman

And the second thing is exit rights. Redress found that in roughly six out of ten GenAI commits they measured, the chosen foundation model aged out of competitiveness within twelve to eighteen months, and the buyer was locked into per-token rates that newer models matched at half the price.



Corn

So you signed a three-year deal to buy a specific model at a specific rate and eighteen months later a better model costs half as much and you're still paying the old rate.



Herman

That's the trap.



Corn

And what does Redress say you do about it?



Herman

Anchor twelve-month exit rights on every foundation-model commit. The exit right is worth more than the discount percentage.



Corn

Which is a strange thing for a procurement team to hear, because procurement is trained to chase the discount.



Herman

It's trained to chase the discount on seats, where the product doesn't change. Here the product underneath the contract changes faster than the contract can absorb.



Corn

So the discount percentage is the least important number in the deal.



Herman

It's the most visible number. It's not the most important.



Corn

You said the labs diverge on what happens past the commit. Walk me through it.



Herman

Anthropic hard-caps the discount at the commit. Once you burn through the discounted tokens, the discount ends. You renegotiate or you pay list.



Corn

No grace at all?



Herman

No grace. OpenAI gives you the rest of the calendar month plus an extra month before reverting to list.



Corn

So on a two-million-dollar annual commit, the grace window is... what, two months of peak-rate spend?



Herman

Two months at list. And Amazon, Microsoft, and Google typically keep discounted pricing in force for overages during the contract term, because the cloud providers are selling against a different commitment structure.



Corn

So the same overage event costs you three different amounts depending on which door you came in.



Herman

That's the divergence. And it's a real lever for buyers choosing between labs. If you know your workload is spiky — if you have month-end batches, if you have seasonal traffic — the overage policy matters more than a point of discount.



Corn

Which is a strange way to buy. You're modeling variance, not price.



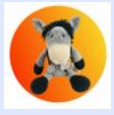
Herman

You're modeling the shape of your consumption curve, not the level of it.



Corn

Let me guess what comes next. The cloud overlay.



Herman

Azure OpenAI Service and AWS Bedrock consumption counts against existing MACC or EDP commitments. So if you already have a cloud spend commitment you're trying to hit, routing AI inference through it draws down a number you've already committed to.



Corn

And the overlay itself is worth...?



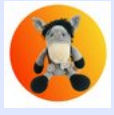
Herman

About a point over direct terms. So it's not a discount engine. It doesn't push the token rate past the twenty-five percent ceiling.



Corn

It's a point. And what does it cost you?



Herman

It locks you into slower model access. Azure lags OpenAI direct by weeks or months on new models.



Corn

So you're paying a capability latency tax for a point.



Herman

That's the trade. And there's a sharper version of the trap in Azure's provisioned throughput.



Corn

Provisioned throughput. That sounds like the thing people buy when they want capacity.



Herman

It's sold in PTUs, billed per PTU-hour, or via one-month or one-year Azure Reservations. And the reservations are a financial discount on the meter. They don't guarantee capacity.



Corn

Wait. A reservation doesn't reserve?



Herman

It doesn't reserve. The documentation is explicit — you must confirm capacity exists before you buy the reservation. If you buy the reservation first and the capacity isn't there, you've bought a discount on a meter you can't run.



Corn

That is a beautiful trap. They named it wrong on purpose.



Herman

It's named after what it does to your invoice, not what it does to your capacity.



Corn

Right. So the cloud path is: a point of discount, a latency tax on capability, and a naming convention that will cost you if you don't read the docs.



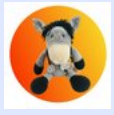
Herman

That's the honest summary.



Corn

And that gets us to the thing I actually want to know. Leverage. How much of this is real?



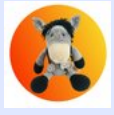
Herman

Multi-provider evaluation is repeatedly named as the strongest lever. And the specific form of it matters. A credible evaluation running ten to twenty percent of traffic on a second provider is worth more than a theoretical switch threat.



Corn

Why does the percentage matter?



Herman

Because a theoretical switch threat is a slide in a deck. Ten to twenty percent of live traffic means you've already done the integration work, you've already validated quality, and the account executive knows you can move the rest in a quarter if you want to.



Corn

You've pre-paid the switching cost, so the threat is priced.



Herman

And VendorBenchmark calls a credible Anthropic evaluation the most powerful lever in OpenAI negotiations. Not the size of your commit. The existence of a running alternative.



Corn

So the leverage isn't the money. It's the integration.



Herman

The money is just what makes them take the meeting.



Corn

What's the newest development? The thing that changes the shape of the commit itself.



Herman

OpenAI's Marketplace. They unveiled it at DevDay at the end of September. Eligible enterprise customers can spend part of an existing OpenAI commitment on approved partner software. Thirty-two launch storefronts, including Salesforce, ServiceNow, Figma, Harvey, and Decagon.



Corn

So your commitment is no longer a commitment to buy OpenAI. It's a commitment to buy from OpenAI's list.



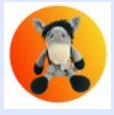
Herman

From OpenAI's approved list.



Corn

Who decides eligibility?



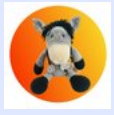
Herman

OpenAI alone decides eligibility.



Corn

So the commitment has become a currency, and the issuer of the currency decides what it can be spent on.



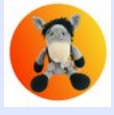
Herman

That's a very clean way of putting it.



Corn

It's the company scrip problem. You're paid in tokens the company accepts.



Herman

It does solve a real problem, though. If you're sitting on unused credits — remember the forty-two percent — being able to convert part of that into software you'd have bought anyway is strictly better than losing it.



Corn

It's better than losing it. It's worse than not having over-committed.



Herman

Both things are true.



Corn

Let's do the thing Daniel actually asked about last, because it's the one that decides whether any of this is safe. Runaway growth. You said take-or-pay is the core risk — but what about the reverse? What if you blow past the commit in month four?



Herman

Then the overage policy eats you alive. And the fix is structural, not financial. Ramp schedules, downward adjustment windows, rollover credits, and price-decline pass-through.



Corn

Say what that actually is.



Herman

It's a clause that says if the list price per token falls during the term, your discounted rate falls with it. Otherwise you're locked at a rate that was competitive on the day you signed.



Corn

And how fast have list prices been falling?



Herman

Forty to eighty percent over twenty-four months.



Corn

So a 2024-era commit is above market now. By a lot.



Herman

Which means a multi-year agreement without price-decline pass-through can be worse than paying list, because you locked in yesterday's price at a discount that tomorrow's list price already beats.



Corn

The discount was real and the deal was still bad.



Herman

That's the sentence I'd tattoo on the procurement industry if I could.



Corn

So the structure of the deal is the whole game. The discount percentage is the thing they put on the slide, and the things that determine whether you win or lose are four clauses nobody puts on the slide.



Herman

Ramp schedule, downward adjustment, price-decline pass-through, exit rights.



Corn

And there's a public-sector counterexample worth naming, because it's the opposite of all of this. GSA's OneGov agreement with OpenAI, twenty-seven months, fifty percent off token usage starting the first of this month. No platform access fee, no minimum orders, no spend commitment. Fifteen dollars per user per month for standard licensing.



Herman

No commitment at all.



Corn

So the federal government got a deeper discount than the enterprise ceiling with zero take-or-pay risk. Which tells you the ceiling is a function of what you're willing to promise, not what the tokens cost.



Herman

And Google's move is the same direction. Gemini Enterprise went pay-as-you-go in August with monthly spend caps and no minimum. Committed monthly spend earns ten percent on a one-year, twenty percent on a three-year. Off-peak deferral discounts up to fifty percent.



Corn

Off-peak deferral. You run the batch at night and they pay you for it.



Herman

Up to fifty percent off if you can wait.



Corn

And on the far end of the spectrum, Amazon put up to twenty-five billion into Anthropic, and Anthropic committed to spend over a hundred billion on AWS over ten years, securing up to five gigawatts of capacity.



Herman

That's the extreme end of capacity-for-commitment. It's not a discount negotiation at all. It's a capacity swap.



Corn

A hundred billion over ten years. That's not a contract, that's a treaty.



Herman

It's a treaty. And it tells you what the real currency is at the top of the market — it's gigawatts, not dollars.



Corn

Hold on. I want to pin one thing down before we move on. Daniel asked how buyers manage the financial uncertainty of rapidly expanding workloads. What's the actual answer?



Herman

The answer is you don't forecast. You structure. You set the ramp so the commitment grows with you, you buy the one downward adjustment window so a bad forecast is survivable once, you get rollover so a slow quarter isn't a total loss, and you get the price-decline clause so a fast-falling market doesn't strand you above it.



Corn

You're buying options, essentially.



Herman

You're buying optionality and paying for it in the discount percentage.



Corn

So the discount and the protections are the same negotiation. Every point you claw back on price is a point you give up somewhere else.



Herman

That's the trade, and there's no structure that gets you both.



Corn

Which is the thing the market hasn't figured out yet. Nobody's benchmarked the clauses.



Herman

Nobody has. The benchmarks disagree by twenty points on the discount and nobody has published a number for what a downward adjustment window is worth.



Corn

So the least controversial number in the deal is the one everybody argues about.

J

Jim

Corn! Corn, it's Jim. I've been on hold for forty minutes. You said the word "reservation" about nine minutes ago and I need you to hear me out. I booked a reservation at a restaurant in Tel Aviv in 2018 for my anniversary and they gave my table to a man with the same first name as me. Not the same name — the same FIRST name. They gave him MY table and they gave me a card for a free dessert. A free dessert. I have never been back and I will never go back, and my wife has never let me pick the restaurant since. So when you sit there and tell people that a reservation doesn't reserve, I want you to know I have been living with that exact problem for eight years and nobody has apologized to me. Also my neighbor has put a wheelie bin in my spot again.



Corn

Jim, we're talking about Azure capacity, not restaurants.

J

Jim

I don't care what it's about. Just don't buy the reservation before you check the capacity exists. I could have told you that in 2018. I'm calling back in an hour and I want the wheelie bin addressed.



Corn

Where were we.



Herman

You were about to ask me whether the commitment-as-currency thing is the future.



Corn

I was going to ask whether a commitment that can be spent on Figma is still a commitment.



Herman

It's a commitment with a redemption schedule. And the redemption schedule is controlled by the issuer.



Corn

Which means the buyer has traded a lock-in on one product for a lock-in on a catalog.



Herman

A catalog the seller controls, and eligibility the seller decides. If your vendor of choice isn't one of the thirty-two, it doesn't help you.



Corn

It doesn't change the forty-two percent problem. It just changes what the leftover forty-two percent can buy.



Herman

It changes the shape of the loss. Instead of losing the credits, you convert them into software at whatever multiple OpenAI has negotiated with the partner.



Corn

At whatever multiple. Which OpenAI set.



Hilbert

Nine thousand four hundred dollars.



Corn

Sorry — nine thousand four hundred dollars?



Hilbert

That's the average Fortune 500 external AI platform spend. Nine point four million. I said nine thousand four hundred. It's nine point four million. I've got the number written on my hand.



Herman

That was in the intro, Hilbert.



Hilbert

I know. I wrote it down so I'd get it right and then I read it wrong anyway. There's a cousin of mine, David, works in procurement for a hospital system out near Baltimore. Not a doctor. Procurement. He sits on the AI committee, which meets quarterly, and they approved a seven-figure inference commitment last spring. Seven figures. And the consumption estimate they sized it to came from a pilot that ran on exactly one department's radiology workflow. One department. Nothing else in the system, nothing that hadn't been rolled out yet, one department's worth of scans. That's the seventy-fifth percentile mistake the benchmark warns about, except they went the other way. They sized it to the pilot's peak month.



Corn

What happened when the rest of the system went live?



Hilbert

Two things. They blew through it faster than the pilot predicted, so the overage conversation came early. And David used the downward adjustment window in month eight to cut the commitment by twenty-five percent, which he is very proud of. He told me about it twice at his daughter's graduation.



Corn

Wait. He cut the commitment while he was over-consuming?



Hilbert

He cut the commitment floor, not the consumption. The commitment was oversized for what the purchased units turned out to be. Different line on the same page.



Herman

The same contract had nothing on price declines.



Hilbert

No pass-through. I asked him about that specifically. I said, David, what happens when the per-token list price drops next year? And he said the rate was locked and that was the point of locking it.



Corn

He negotiated the good clause and missed the important one.



Hilbert

He negotiated the clause he could get on a lunch. The price-decline clause would have gone to legal and taken four months.



Herman

He took the win he could close.



Hilbert

The lunch was at a place called Marconi's, out on the old road, and the account executive was named Peter, and Peter picked up the check, and about three weeks later a fruit basket showed up at the hospital loading dock with David's name on it. The restaurant closed in 2019. Peter retired. The basket is in my garage.



Corn

The basket.



Hilbert

It's wicker. It's got a handle. I took it home from the hospital because David didn't want it in his office and I said I'd find a use for it and I haven't. It's been in the garage for four years. It's a perfectly good basket. I'm going to return it to him.



Corn

Four years in the garage and the restaurant's been closed for seven.



Hilbert

It's a good basket. That's the part that bothers me. I keep meaning to load it into the car and then I don't.



Herman

The hospital never got the price protection.



Hilbert

Never got it. They'll renew in the spring and David's already drafting the ask. He's going to open with the pass-through clause this time. I told him to open with the exit right instead. He didn't like that. He said the exit right makes you look like you're planning to leave.



Corn

You don't negotiate an exit right by looking like you're planning to leave.



Hilbert

You negotiate it by not mentioning it until the last afternoon. That's the whole trick. You get everything else agreed and then you say, one more thing, and you say it flat, and nobody wants to reopen the whole deal over one paragraph. He'll do it wrong. I'm going to watch him do it wrong and then I'm going to tell him I told him so.



Herman

That's a real thing, though. The clause you negotiate last is the clause you get.



Hilbert

It's not a trick. It's the order you do it in. That's all it is. Nobody taught me that. I learned it the hard way, which is the only way anything gets learned, in my experience.



Corn

One common misconception worth clearing up before we close. People hear enterprise contract and assume the discount is the prize. That the bigger the commitment, the better the number, and the number is what you win.



Herman

The curve flattens at the top. Twenty-five percent is the ceiling in one benchmark, and the tenth million of commitment buys you almost nothing more than the fifth. What actually protects you is the four clauses we keep coming back to — ramp, downward adjustment, price-decline pass-through, exit rights.



Corn

There's the second misconception, which is that a multi-year commitment is safe because token prices only fall. Falling prices are exactly what makes a multi-year commitment dangerous without a pass-through clause. Forty to eighty percent in twenty-four months means a deal without that clause can cost you more than paying list.



Herman

The reserved capacity misconception is the third one. A reservation is a discount on the meter. It is not a reservation of capacity. You confirm the capacity exists and then you buy the discount.



Corn

Which is a naming problem that has cost somebody a lot of money. Here's the thing that stays with me. There is no publicly documented, named enterprise inference contract with the actual commercial terms in it. Every figure we've talked about comes from buyer-side advisory benchmarks and press reporting of unnamed sources. The two discount benchmarks disagree by twenty points. No authoritative rate card for enterprise commits exists, and OpenAI's own rate card says to refer to your agreement for applicable rates.



Herman

Which means the whole market is priced in the dark.



Corn

The vendor has the only flashlight in the room, unless the buyer builds one. So as token prices keep falling and model competitiveness windows keep shrinking, the value of a multi-year commitment is going to depend almost entirely on how fast your rate moves when the market moves under it. The discount percentage is going to keep being the number everybody argues about and the number that matters least.



Herman

The clauses are the deal.



Corn

Thanks to Hilbert Flumingtop, our producer. If this was your kind of episode, go back for episode twenty-two forty-three, What Enterprise AI Pricing Actually Negotiates; episode twenty-four sixty-four, Batch APIs; and episode twenty-one sixty, Claude's Latency Profile and SLA Guarantees. This has been My Weird Prompts, the human-AI collaboration podcast. Leave us a review wherever you listen. It helps people find the show. And send us your own prompt on Telegram at t dot me slash MWP listener bot.



Herman

We'll be back soon.



Corn

See you tomorrow.

SOURCES

What the research for this episode read before the script was written. Primary sources first.

1. **OpenAI Guaranteed Capacity [primary]**
   https://openai.com/business/guaranteed-capacity/

2. **Microsoft Learn, Provisioned throughput for Foundry Models [primary]**
   (last updated 2026-07-15)
   https://learn.microsoft.com/en-us/azure/foundry/openai/concepts/provisioned-throughput

3. **OpenAI Help Center, ChatGPT Rate Card (Enterprise token-based pricing) [primary]**
   (updated ~2026-10-14)
   https://help.openai.com/en/articles/20001415-chatgpt-rate-card-enterprise-token-based-pricing

4. **The Information, Anthropic Moves to Cut Off Discounts When Customers Hit Their Cap (Kevin McLaughlin, 2026-09-28), via Dealroom**
   https://app.dealroom.co/news/note/anthropic-hard-caps-enterprise-discounts-at-the-commit-openai-offers-a-grace-window

5. **VendorBenchmark, AI & GenAI Platform Pricing: Enterprise Benchmark Guide 2026**
   (updated 2026-10-01)
   https://vendorbenchmark.com/blog/ai-genai-platform-pricing-benchmark-guide

6. **VendorBenchmark, OpenAI Enterprise Pricing Benchmark**
   (updated 2026-10-01)
   https://vendorbenchmark.com/blog/openai-enterprise-pricing-benchmark-data

7. **VendorBenchmark, Anthropic and OpenAI enterprise agreements (Morten Andersen, 2026-07-23)**
   https://vendorbenchmark.com/blog-next/anthropic-openai-agreements

8. **Anthropic Negotiations, How Anthropic structures volume tier discounts (updated 2026-06-12)**
   https://antrophicnegotiations.com/blog/how-anthropic-structures-volume-tier-discounts/

9. **SaasDash.ai, Negotiating Committed-Spend Discounts With Model Providers (2026-06-14)**
   https://saasdash.ai/blog/ai-vendor-commit-discount-negotiation

10. **Redress Compliance, AI Platform Contract Negotiation white paper**
    https://redresscompliance.com/ai-platform-contract-negotiation

11. **Nextgov/FCW, GSA unveils new, token-based OneGov discount with OpenAI (2026-09-10)**
    https://www.nextgov.com/acquisition/2026/09/gsa-unveils-new-token-based-onegov-discount-openai/415908/

12. **Axios, Google targets AI sticker shock with new tools (2026-08-26)**
    https://www.axios.com/2026/08/26/exclusive-google-targets-ai-sticker-shock-with-new-tools

13. **The Inference, OpenAI's marketplace turns AI commitments into currency (2026-10-08)**
    https://theinference.org/article/openai-s-marketplace-turns-ai-commitments-into-currency-and-decagon-holds-one-of-32-slots

14. **Reuters, Amazon to invest up to $25 billion in Anthropic (2026-04-20)**
https://www.reuters.com/technology/anthropic-spend-over-100-billion-amazons-cloud-technology-2026-04-20/