

Copilot Credits

The September 1 Exposure

Messaging, communications and field enablement response to the usage based billing transition.

37%

Drop in Business allowance

44%

Drop in Enterprise allowance

DECISION REQUIRED

1 September 2026

TO

Copilot GTM leadership

FROM

Naveen Raj Kanagaraj, independent

DATED

18 August 2026, rev. 16 Sep

Spec recommendation

Independent work, not affiliated with GitHub or Microsoft. Diagnosis, repositioning, communications and field enablement.

*

2026

The transition landed. The invoice has not.

1 JUNE

Premium requests replaced by metered AI Credits across every plan.

PRICE

Base plans unchanged. Business \$19, Enterprise \$39 per user, per month.

SCOPE

Completions stay free. Chat, CLI, agents and code review are metered.

COVER

Promotional credits end 1 September. Sonnet 5 promotional token pricing ends 31 August.

SIGNAL

Public backlash is led by individual developers. Enterprise exposure is deferred, not avoided.

We did not raise prices. We removed the ability to forecast them.

A COST INCREASE

is absorbed.

Negotiated at renewal. It stays inside engineering.

A FORECASTING FAILURE

is escalated.

It reaches finance, triggers procurement review and a spend freeze.

**We produced the second and are being criticised for the first.
Every fix that adjusts price solves the wrong problem.**

Credits burn on tokens at each model's published rate. Output costs \$5 per million tokens on Claude Haiku 4.5 and \$25 on Claude Opus, and for code review the model is chosen automatically and not disclosed.

The safety net came out in the same release as the meter.

BEFORE

Allowance runs out.
Model falls back.
Work continues, degraded.

Graceful. No decision required.

AFTER

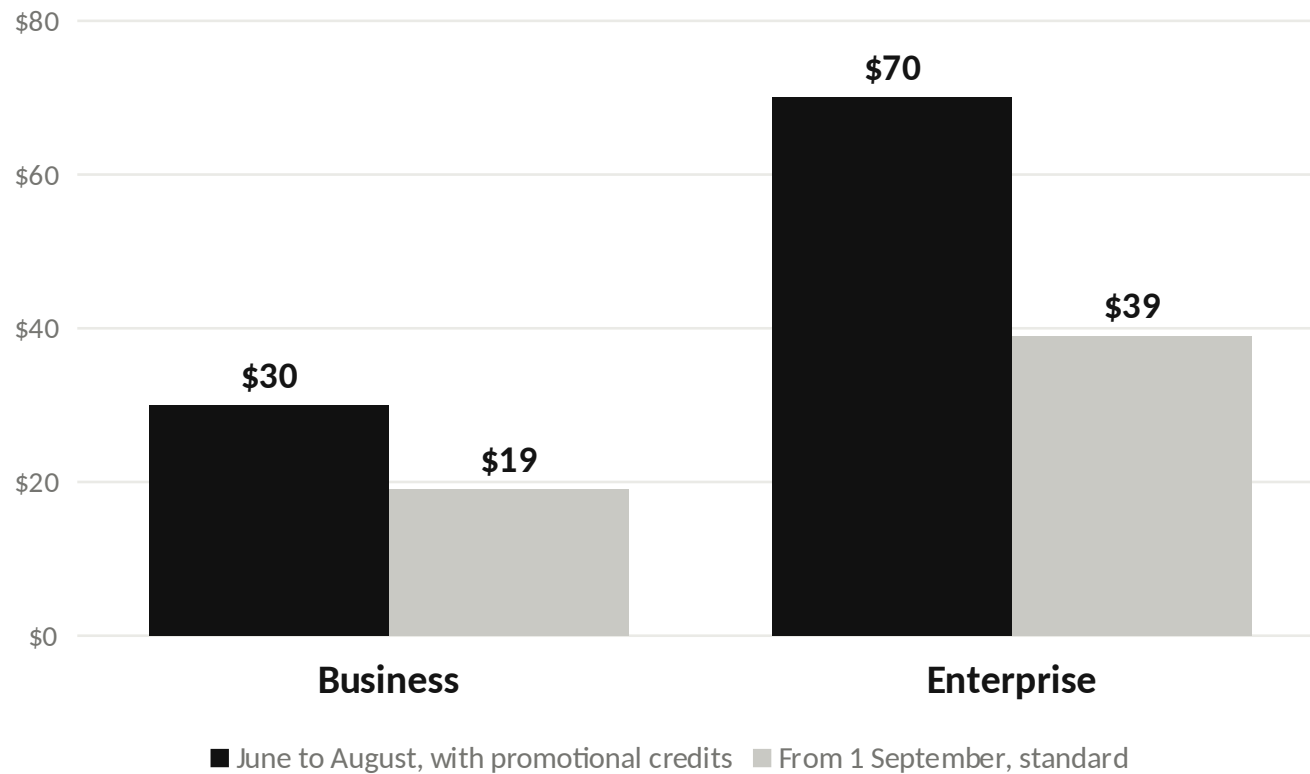
Allowance runs out.
No fallback exists.
Overage by default, or work stops.

Binary. Overage unless an admin opts out.

THE CONSEQUENCE WE HAVE NOT MESSAGED

The field is selling budget caps as the answer to unpredictability. The cap is also an off switch, and overage stays on unless an admin disables it. A user budget can block a developer while the pool still has credits.

On 1 September, included usage falls by up to 44%.



BUSINESS

37% less

3,000 to 1,900 credits (\$30 to \$19) per user

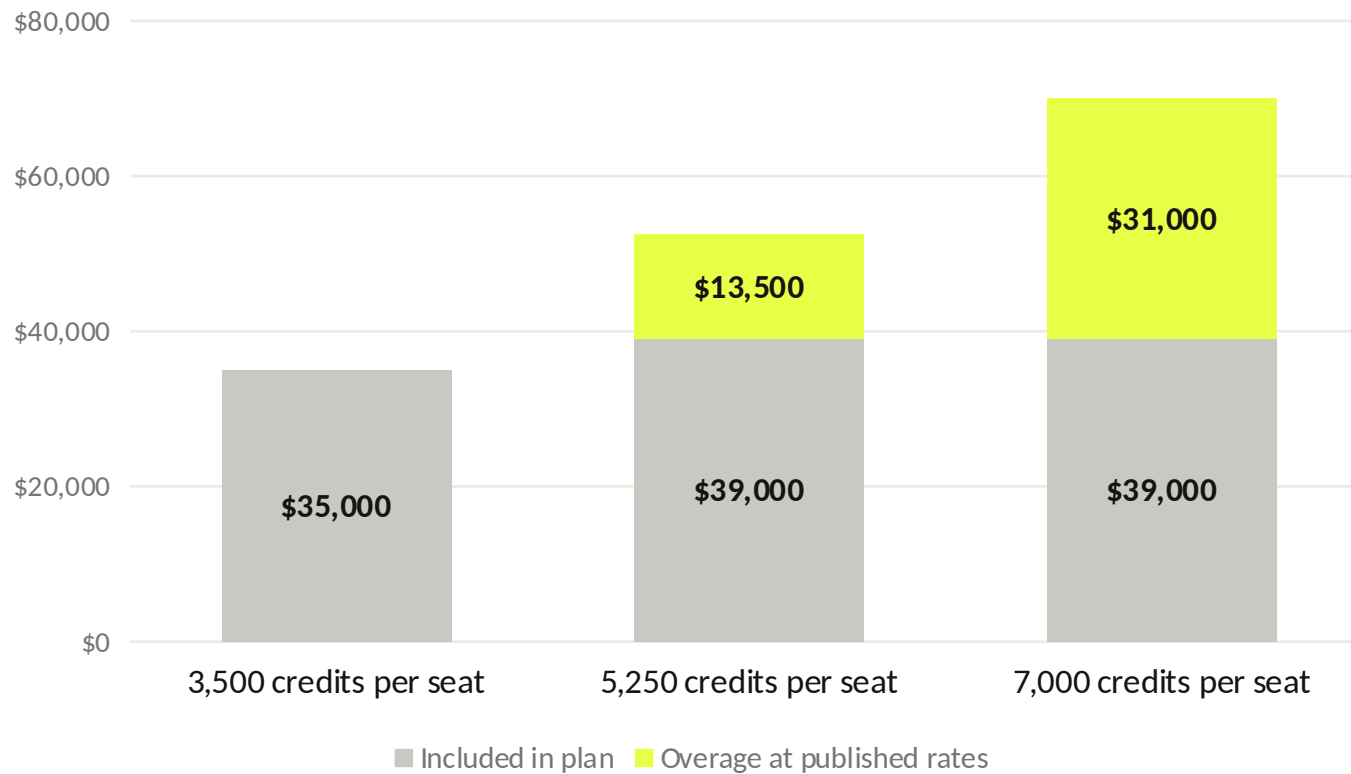
ENTERPRISE

44% less

7,000 to 3,900 credits (\$70 to \$39) per user

The promotional amount replaced the standard allowance for three months. It was not added on top. Seat prices do not change.

What 1,000 Enterprise seats pay in September.



AT FULL PROMOTIONAL USE

\$31K

of monthly overage, billed by default, if usage holds at the promotional allowance.

READ THIS AS A RANGE

Below 3,900 credits per seat, September adds no overage. Above it, every credit is billed. The customer cannot see in advance which side of that line they land on.

Pooled average consumption per seat at 1 credit = \$0.01, against a 3,900,000 credit pool. Directional, not a forecast.

From predictable seat to governed platform.

WAS

"One price per developer.
Stop thinking about it."

IS

**"Spend you can see, attribute
and govern before it happens."**

POSITIONING STATEMENT / PLATFORM ENGINEERING LEAD

For the platform lead who now owns a variable AI budget, Copilot meters agentic work against a pooled organisation allowance with budgets at enterprise, cost centre and user level, so spend is attributable, can be capped, and holds up in a budget review.

What it deliberately does not claim: that the bill is predictable. It is not, and claiming otherwise breaks trust a second time.

One reposition. Three translations.

DEVELOPER

Sees the meter

"See the cost before you commit."

PROOF

Completions stay free and unmetered.
Every model's per-token rate is published.

CHANNEL

In product, docs, changelog

PLATFORM LEAD

Owns allocation

"Allocate by cost centre, not by seat."

PROOF

Pooled usage removes stranded capacity.
Budgets set at enterprise, cost centre or user level.

CHANNEL

Admin console, CSM, webinar

FINANCE

Sees the variance

"You choose the ceiling in advance."

PROOF

On exhaustion: continue at published rates, or cap. Overage is the default.

CHANNEL

Procurement FAQ, renewal brief

Everyone has this problem. Not everyone shows the customer.

Agentic coding consumes frontier inference. That cost is real for every vendor. The only question is who absorbs the variance, and for how long.

FLAT RATE SEAT

The easier sell

Predictable invoice. Easy to buy.

The vendor absorbs variance, then rations quietly through throttles, fair use limits and downgrades at peak. Spend cannot be attributed to a team, because there is no spend to attribute.

METERED AND GOVERNED

Copilot, from 1 June 2026

Spend is attributable and capped in advance.

The invoice moves. The trust cost is paid up front and in public, which is where we are standing now.

Speculation: flat rate agentic pricing does not survive current inference economics. Vendors still selling it will meter, and their customers will arrive with no allocation model at all.

Land the cliff before it lands on them.

19 AUG	Admin notification	Every admin receives their own June to August consumption beside their standard September allowance.
19 AUG	Field briefed first	AEs and CSMs get the cap trade off before customers get the email.
25 AUG	Guided configuration	CSM led budget setup, ranked by projected gap rather than by ARR.
1 SEP	Say it plainly, in public	Changelog and docs state that promotional credits have ended. Silence converts a billing change into a trust story.
EARLY OCT	First unsubsidised invoice	Outreach to every account above forecast before they open it.

The objection that decides Q4 renewals.

"We cannot budget this. The variance is too wide."

WHAT WE SAY NOW

"Set a cap. Then you will never see an unplanned overage."

True, and incomplete. It omits what happens at the cap. An admin who sets one on our advice and finds their team blocked in week three will not call that a configuration issue.

WHAT TO SAY INSTEAD

"You are choosing between a spend ceiling and a throughput ceiling."

- Size the cap against observed June to August usage, not the standard allowance
- Prefer cost centre budgets: a user budget blocks that developer even when the pool has credits
- Alert well below the ceiling so the admin decides, not the meter
- Run the projection sheet with them, using their own actuals, not a worked example

One ask. Marketing cannot close this alone.

Budget controls, pooled usage and the preview bill all shipped. This is not a request for tooling we lack.

THE ASK

Restore a degraded capability floor at the cap.

When the pool is exhausted and the organisation has chosen to cap, route to a low cost model at reduced concurrency instead of stopping. The admin keeps a hard spend ceiling. The developer keeps a working tool.

WHY THIS AND NOT A DISCOUNT

A taper postpones the conversation by one cycle and costs revenue in every cycle after. A floor is paid for once.

IF THIS IS DECLINED

The claim weakens from governed to attributable. Marketing's fallback is alerting well ahead of the ceiling and a named approval path, which manages the stop rather than preventing it.

What this costs us.

WE GIVE UP

The predictability message.

Competitors selling flat rate run it against us for at least two quarters and it works. We should expect to lose price sensitive mid market deals, and should not pretend otherwise in the forecast.

WE KEEP

A claim that stays true.

Attribution and governance can be demonstrated in a console today and will not be retracted. The alternative fails on the first spiky month and costs the account, not just the deal.

THE BET

That in enterprise software, trust compounds and price advantages do not. We are absorbing the reaction in public, and the allocation layer becomes table stakes as the rest of the category meters. If that reading is wrong, this is the wrong strategy, and the next slide says so by November.

Four numbers, and what would prove me wrong.

01**Budget configuration rate**

Organisations with a budget set at any level, by 30 September.

02**Cap proximity in week two**

Share of organisations whose usage approaches their own ceiling within fourteen days. The only signal that arrives early enough to recalibrate before the invoice.

03**Forecast gap at first invoice**

Admin projected against actual October spend. The direct measure, and the number I would be held to.

04**Billing cited in Q4 churn**

Non renewals naming predictability. Lagging and contaminated. Read alongside 03, never alone.

Falsification: If the October gap is wide and configuration is high, governance did not fix forecasting. The recommendation was wrong and should be withdrawn, not iterated.

Tell them before the invoice does.

Reposition from predictable seat to governed platform, name the cap trade off before customers discover it, and ask Product for a capability floor so governance does not mean stoppage. The promotional window ends in two weeks either way.

01

Say it publicly on 1 September

02

Brief the field before the customer

03

Ship a floor at the cap

Decision required by 1 September 2026.

Working, sources and assets.

A	Assumptions and sources	Every figure traced to GitHub's announcement and billing docs, with derivations.
B	Account risk segmentation	Ranking logic for outreach: projected gap, not ARR.
C	Full messaging matrix	Proof points, disqualifiers and phrases to avoid.
D	Admin notification email	Draft copy for the 19 August send.
E	Objection one pager	Cap trade off briefing for AEs and CSMs.
F	Spend projection sheet	CSM calculator: an account's own actuals against its September allowance.
G	Research basis and limits	What the evidence is and what it cannot support.

Assumptions and sources.

Business plan	\$19 per user per month, including \$19 in monthly AI Credits	GitHub
Enterprise plan	\$39 per user per month, including \$39 in monthly AI Credits	GitHub
Promotional credits	Replace the standard amount: 3,000 and 7,000 credits per user, 1 June to 1 September	GitHub Docs
Unmetered	Code completions and Next Edit suggestions, all plans	GitHub
Fallback	Discontinued. Overage enabled by default for organisations	GitHub Docs
Budget controls	Enterprise, cost centre and user level. Included usage pooled	GitHub
Sensitivity model	1,000 Enterprise seats at 3,500, 5,250 and 7,000 credits against a 3,900,000 pool	Derived
Not verified	September consumption, budget adoption, competitor pricing	Unknown