KI & COPILOT

Copilot: Modelle nach Nutzungslimit nur noch gegen Copilot Credits (kostenpflichtig)

Consulting Briefing · 21.11.2026 · boddenberg.de

Executive Summary

Copilot bekommt einen Münzeinwurf, und der Schlitz sitzt direkt neben dem Prompt-Feld. Laut Microsoft-Ankündigung MC1557754 können Nutzer mit einer Microsoft-365-Copilot-Lizenz ab Dezember 2026 weiter mit ihrem gewählten Premium-Modell arbeiten, wenn das in der Lizenz enthaltene Nutzungslimit aufgebraucht ist. Weiter geht es dann über nutzungsbasierte Abrechnung, also gegen Copilot Credits. Wer nicht zahlen will oder darf, wird auf Auto umgeleitet, das Microsoft weiterhin ohne Credit-Verbrauch anbietet. Betroffen sind Copilot Chat in der Microsoft-Copilot-App sowie Copilot in Word, Excel und PowerPoint, im Web, auf dem Desktop und mobil.

Die Uhr tickt laut: Microsoft hat die Meldung am 20.11.2026 aktualisiert. Fassung 2 nennt jetzt Anfang Dezember für das Frontier-Programm und Mitte Dezember für die allgemeine Verfügbarkeit in der weltweiten Cloud. Das sind weniger als 30 Tage. Und es gibt eine Standardeinstellung, die für dich entscheidet, wenn du es nicht tust: Spending Policies mit aktivem „Auto-apply new services“ nehmen den neuen Dienst „Models with usage-based billing in Chat and Office“ automatisch auf. Laut Microsoft Learn ist diese Option bei jeder neuen Policy vorausgewählt.

Fakten: Das Wichtigste auf einen Blick

Quelle: Microsoft-Ankündigung MC1557754, zuletzt geändert am 20.11.2026. Vorläufer: MC1479276 (neues Preismodell) und MC1519122 (Modelle nur über nutzungsbasierte Abrechnung).

Frontier: Anfang Dezember 2026. Allgemeine Verfügbarkeit: Mitte Dezember 2026. Meldung läuft bis 28.02.2027.

Nach dem Limit: Auto ohne Credits oder gewähltes Modell gegen Credits, Letzteres nur mit passender Spending Policy.

Neu in der Oberfläche: Credit-Symbol (ausgefüllt = kostet, leer = kostet nicht), Banner nach Erreichen des Limits, Befehl /cost, ausgegraute Modelle im Model Picker.

Admin-Ort: Microsoft 365 Admin Center › Copilot › Cost Management › Configuration.

Worum geht es im Detail?

Zum Hintergrund: Microsoft hat im September das Copilot-Preismodell umgebaut. Im Tech Community Blog erklärt der Konzern die Trennung in „Everyday AI“ und „Advanced AI“. Alltags-KI steckt in der Copilot-Lizenz pro Benutzer und Monat, dazu gehören nach Microsofts Angaben GPT-5.6 und Sonnet 5, neue Modelle wie Opus 5 sind „mit Limits“ enthalten. Alles, was lange läuft oder richtig Rechenleistung frisst, also Cowork, Code, Autopilot und Frontier-Modelle, wird über Copilot Credits abgerechnet. Die Lizenz ist dabei Eintrittskarte, nicht Flatrate. Das ist ehrlich

gesagt das Geschäftsmodell eines Freizeitparks: Eintritt zahlen, und für die gute Achterbahn gibt es Chips an der Kasse.

MC1557754 schließt jetzt die Lücke zwischen beiden Welten. Bisher war beim Erreichen des Limits einfach Schluss mit dem Premium-Modell. Ab Dezember erscheint stattdessen ein Banner mit einer Wahl: weiter mit Auto, kostenlos, oder weiter mit dem gewählten Modell, kostenpflichtig. Auto ist dabei kein Billigmodell, sondern ein Router, der laut Microsoft bei jeder Anfrage Genauigkeit, Geschwindigkeit und Kosten abwägt und Modell und Denktiefe passend wählt. Für 90 Prozent aller Mails, Zusammenfassungen und Excel-Formeln reicht das völlig. Für die restlichen zehn Prozent hält jemand in deinem Unternehmen sein Lieblingsmodell für unverzichtbar und wird jetzt gefragt, ob er dafür Geld ausgeben möchte. Spoiler: Es ist ja nicht sein Geld.



Abbildung 1: Ein Prompt nach dem Limit. Entscheidend ist nicht der Klick des Nutzers, sondern ob eine Spending Policy den neuen Dienst enthält.

Was genau sich ändert, lässt sich in vier Punkten fassen. Erstens taucht in den Spending Policies ein neuer Dienst auf, „Models with usage-based billing in Chat and Office“. Zweitens bekommt die Oberfläche ein Credit-Symbol: ausgefüllt heißt, der Zähler läuft, leer heißt, er läuft nicht. Ein Mouseover zeigt Details, und wer /cost als Prompt abschickt, sieht die Kosten der aktuellen Sitzung. Drittens zeigt der Model Picker künftig alle Modelle, die in Chat sowie in Word, Excel und PowerPoint angeboten werden, auch die, auf die der Nutzer keinen Zugriff hat. Die sind ausgegraut. Viertens greift Auto-apply: Bestehende Policies mit dieser Option übernehmen den neuen Dienst ohne weiteres Zutun.

Warnung: Der Cowork-Pilot als trojanisches Pferd

Klassiker aus den letzten Wochen: Ein Unternehmen richtet für den Cowork-Pilot schnell eine Spending Policy ein, Bereich „Alle Benutzer“, kein Nutzerlimit, Auto-apply an, weil das eben vorausgewählt war. Ab Mitte Dezember gilt diese Policy auch für Premium-Modelle in Chat und Office. Aus dem Pilot mit zwölf Leuten wird eine offene Kreditlinie für die gesamte Belegschaft. Niemand hat etwas falsch gemacht, alle haben nur auf „Weiter“ geklickt. So beginnen die besten Controlling-Geschichten.

Die zweite Mechanik, die du kennen musst: Ist ein Nutzer Mitglied in mehreren Policies, werden die Einstellungen nicht zusammengeführt. Laut Microsoft Learn gewinnt die Policy mit dem höchsten Nutzerlimit, danach die mit dem größten Gesamtlimit, danach die jüngste. Wer also eine großzügige Policy für die Entwicklungsabteilung und eine knappe für alle anderen anlegt, sollte sehr genau wissen, wer in welcher Gruppe steckt. Und wenn das Gesamtlimit einer Policy erreicht ist, verlieren alle Nutzer dieser Policy den Zugang zu sämtlichen enthaltenen Diensten bis zum Monatsersten. Ein einzelner Vielnutzer ohne Nutzerlimit kann so am 20. des Monats den Cowork-Zugang für die ganze Abteilung beerdigen.

Was sind Chancen? Was sind Risiken?

Die Chance ist echt, man muss sie nur wollen. Bisher lief die Diskussion über Premium-Modelle nach dem Muster „alle oder keiner“: Entweder bekommt jeder eine teure Zusatzlizenz, oder die Juristin mit dem 80-seitigen Vertragsentwurf muss mit dem gleichen Modell arbeiten wie der Kollege, der sich Geburtstagsmails formulieren lässt. Mit nutzungsbasierter Abrechnung bezahlst du genau dort, wo das bessere Modell einen Unterschied macht. Beispiel aus der Praxis: Ein Ingenieurbüro mit 120 Leuten hat drei Projektleiter, die regelmäßig Ausschreibungsunterlagen von mehreren hundert Seiten in Word gegen Leistungsverzeichnisse prüfen. Genau diese drei bekommen eine eigene Policy mit Nutzerlimit, alle anderen arbeiten mit Auto. Das kostet überschaubar und erspart die Diskussion, warum der Azubi nicht auch Opus braucht.

Die Risiken sind weniger technisch als organisatorisch. Erstens Kostenkontrolle: Microsoft empfiehlt in der Learn-Dokumentation für Policies ein Nutzerbudget von mindestens 2.000 Credits pro Person und Monat und weist ausdrücklich darauf hin, ein Nutzerlimit zu setzen, damit nicht eine Person alle Credits verbraucht. Einen Preis pro Credit nennt die Learn-Dokumentation nicht, der hängt an Vertrag, Pay-as-you-go-Tarif oder vorab gekauften Kontingenten. Rechne deshalb mit deinem eigenen Vertragspreis, nicht mit Zahlen aus Foren. Zweitens Transparenz: Die Nutzer entscheiden selbst, ob sie weiter zahlen. Ein Banner mit „Weiter“ wird geklickt, das ist so sicher wie die Mittagspause.

Ausgangslage im Tenant	Was ab Mitte Dezember passiert	Kostenrisiko
Keine Spending Policy vorhanden	Nach dem Limit nur Auto, Premium-Modelle ausgegraut	keins, aber Frust und Helpdesk-Tickets
Policy mit Auto-apply an, alle Benutzer, kein Nutzerlimit	Jeder Copilot-Nutzer kann nach dem Limit Credits verbrauchen	hoch, gedeckelt nur durch Gesamtlimit

Ausgangslage im Tenant	Was ab Mitte Dezember passiert	Kostenrisiko
Policy mit Auto-apply an, Gruppe, Nutzerlimit	Nur die Gruppe zahlt weiter, jeder bis zu seinem Limit	kalkulierbar
Policy mit Auto-apply aus	Neuer Dienst nicht enthalten, bis ein Admin ihn hinzufügt	keins, bewusste Freigabe nötig
Nutzer in mehreren Policies	Policy mit höchstem Nutzerlimit gilt vollständig	Überraschung programmiert

Für deutsche Organisationen kommen drei Dinge dazu. Datenschutz: Premium-Modelle stammen von verschiedenen Anbietern, darunter Anthropic. Prüfe für jedes Modell, das du freigibst, ob es in deinem Verarbeitungsverzeichnis und deiner Datenschutz-Folgenabschätzung abgedeckt ist und wie es um EU-Datengrenze und Unterauftragsverarbeitung steht. Mitbestimmung: Das Cost-Management-Dashboard zeigt im Bereich Consumption den Credit-Verbrauch je Nutzer und Tag. Das ist eine technische Einrichtung, die Verhalten auswerten kann, also § 87 Abs. 1 Nr. 6 BetrVG, im öffentlichen Dienst das Personalvertretungsrecht. Und Haushaltsrecht: Variable Verbrauchskosten ohne Obergrenze passen in keinen Haushaltstitel. Für öffentliche Auftraggeber ist ein hartes Gesamtlimit keine Empfehlung, sondern Pflicht.

Wichtig: Der Verbrauchsbericht ist ein Leistungsbericht mit anderem Namen

Wer sehen kann, dass Frau Meier im November 4.800 Credits und Herr Schulz 30 verbraucht hat, sieht indirekt, wer wie intensiv mit KI arbeitet. Regle in der Betriebs- oder Dienstvereinbarung, wer diese Ansicht öffnen darf (die Rolle AI Reader reicht dafür), zu welchem Zweck und ob Auswertungen nur aggregiert erfolgen. Sonst wird aus der Kostenkontrolle im ersten Konflikt eine Leistungskontrolle, und dann sitzt der Betriebsrat schneller im Büro der Geschäftsführung als der Copilot eine Zusammenfassung schreibt.

Was müssen wir jetzt schon vorbereiten?

Kurz gesagt: entscheiden, bevor die Standardeinstellung es für dich tut. Die Steuerung läuft komplett über die Spending Policies im Microsoft 365 Admin Center unter Copilot › Cost Management › Configuration. Anlegen und Abrechnungsmethoden ändern dürfen Global Administrator und Billing Administrator, bearbeiten auch AI Administrator und License Administrator. Eigene PowerShell-Cmdlets dafür dokumentiert Microsoft derzeit nicht; für Automatisierung gibt es eine Graph-API für Cost Management. Ein globaler Schalter „nutzungsbasierte Modelle aus“ existiert nicht. Abschalten heißt: den Dienst aus allen Policies entfernen und Auto-apply deaktivieren, damit er nicht über die Hintertür zurückkommt.



Abbildung 2: Frontier-Start, GA und Meldungsende stammen aus MC1557754. Die Zwischenschritte sind unsere Empfehlung.

Wann	Wer	Was konkret
Bis 28.11.2026	Copilot- bzw. Billing-Admin	Alle Spending Policies öffnen und dokumentieren: Bereich (alle Benutzer oder Gruppen), Auto-apply an oder aus, Gesamtlimit, Nutzerlimit, Alarme, Abrechnungsmethode. Mehrfachmitgliedschaften in Gruppen prüfen.
Bis 30.11.2026	Copilot-Admin	Vor dem Frontier-Start: In allen Policies, die keine Premium-Modelle abdecken sollen, Auto-apply ausschalten und den Dienst „Models with usage-based billing in Chat and Office“ entfernen, sobald er auftaucht.
Bis 05.12.2026	Entscheider, Controlling	Festlegen, wer weiter zahlen darf: eigene Sicherheitsgruppe, eigene Policy, Nutzerlimit, Gesamtlimit, Alarm bei 50 und 80 Prozent. Öffentliche Auftraggeber: Obergrenze passend zum Haushaltstitel.
Bis 12.12.2026	IT, Datenschutz, Betriebs- oder Personalrat	Modelle im Verarbeitungsverzeichnis prüfen, Zugriff auf Verbrauchsberichte regeln, Belegschaft und Helpdesk informieren: Banner, Credit-Symbol, /cost, ausgegraute Modelle.
Ab GA Mitte Dezember	Copilot-Admin	Wöchentlich das Consumption-Dashboard prüfen, Anfragen nach mehr Credits über die Seite Credit requests bearbeiten.

Wann	Wer	Was konkret
Januar und Februar 2027	Controlling	Die ersten beiden Abrechnungen auswerten, Limits nachsteuern, bis zum Meldungsende am 28.02.2027 eine dauerhafte Regel festschreiben.

Tipp: Drei Sätze für das Intranet

„Wenn dein Copilot-Kontingent erschöpft ist, wähle Auto, das ist kostenlos und für fast alles gut genug. Ein ausgefülltes Münzsymbol bedeutet, dass dieser Prompt Geld kostet, /cost zeigt dir, wie viel. Ausgegraute Modelle sind kein Fehler, sondern eine bewusste Entscheidung, Tickets dazu bitte mit Begründung an die IT.“ Das spart dir im Dezember geschätzt die Hälfte aller Helpdesk-Anrufe, und die andere Hälfte kommt ohnehin von Leuten, die das Intranet nicht lesen.

Die Einrichtung im Detail beschreibt Microsoft in den Learn-Artikeln Nutzungsbasierte Abrechnung für Copilot Credits einrichten und Spending Policies und Dienstzugriff verwalten. Für die laufende Kontrolle gibt es zusätzlich den Artikel zum Überwachen der Ausgaben für Copilot Credits.

Weiterlesen auf boddenberg.de

› Copilot Managed Runtime wird allgemein verfügbar; App-Erstellung in Copilot Cowork
› Copilot Notebooks wird zu Copilot Workspaces mit Aktivitätsfeed und Teampräsenz

Häufig gestellte Fragen

Was passiert in Microsoft 365 Copilot, wenn das Nutzungslimit für ein Modell erreicht ist?

Ab Dezember 2026 können Nutzer nach dem Limit kostenlos mit Auto weiterarbeiten, das automatisch ein passendes Modell wählt. Nur wer über eine Spending Policy Zugriff auf Copilot Credits hat, kann das gewählte Premium-Modell weiter nutzen, dann gegen Credits.

Wie verhindere ich, dass Mitarbeiter in Copilot Chat Copilot Credits verbrauchen?

Entferne im Microsoft 365 Admin Center unter Copilot › Cost Management › Configuration den Dienst „Models with usage-based billing in Chat and Office“ aus allen Spending Policies und schalte dort „Auto-apply new services“ aus. Gibt es gar keine Spending Policy, können Nutzer ohnehin keine Credits verbrauchen.

Was bedeutet „Auto-apply new services“ bei Copilot Spending Policies?

Die Option sorgt dafür, dass neu unterstützte Copilot-Dienste automatisch in eine bestehende Spending Policy aufgenommen werden. Sie ist standardmäßig aktiviert, weshalb Policies, die etwa

für Cowork angelegt wurden, ab Dezember 2026 auch kostenpflichtige Modelle in Chat und Office freigeben.

Wofür steht das Münzsymbol in Copilot und was zeigt der Befehl /cost?

Ein ausgefülltes Credit-Symbol bedeutet, dass die aktuelle Nutzung Copilot Credits verbraucht, ein leeres, dass sie über die Lizenz abgedeckt ist. Mit dem Prompt /cost zeigt Copilot die Kosten der aktuellen Sitzung an.

Welche Spending Policy gilt, wenn ein Nutzer in mehreren Gruppen ist?

Microsoft wendet genau eine Policy vollständig an, Einstellungen werden nicht kombiniert. Den Vorrang hat die Policy mit dem höchsten Nutzerlimit, danach die mit dem größten Gesamtlimit und zuletzt die zuletzt erstellte.