

Agent 365: Kostenverwaltung soll im Oktober auch Copilot-Studio-Agenten abdecken

Die Zeit der anonymen Bot-Zeche ist vorbei. Ab Oktober bekommt jeder Maker-Agent eine Rechnungsadresse, und irgendjemand in deinem Haus bekommt die Mahnung.

Executive Summary

Microsoft baut die Kostenverwaltung in Agent 365 aus. Bisher hat sie nur zwei Verbraucher im Blick: Copilot Cowork und die Work IQ APIs. Mit der Ankündigung vom 25. September 2026 kommen **Code** und die **Copilot Managed Runtime** hinzu, und für **Oktober** ist die Unterstützung für Agenten aus **Copilot Studio** geplant. Damit landet genau die Agentenklasse in der zentralen Budgetsteuerung, die in den meisten Organisationen am wildesten wuchert: die Bots, die Fachbereiche selbst zusammenklicken.

Warum das wichtig ist? Bislang lebte die Abrechnung von Copilot-Studio-Agenten in einer Parallelwelt. Im Power Platform Admin Center hängst du einzelne Umgebungen per Abrechnungsrichtlinie an ein Azure-Abonnement oder verteilst Prepaid-Kapazität. Cowork und Co. steuerst du dagegen im Microsoft 365 Admin Center über Ausgabenrichtlinien mit Monatsbudgets pro Gruppe und pro Nutzer. Zwei Kassen, zwei Teams, zwei Wahrheiten. Ab Oktober soll daraus eine Kasse werden.

Die Konsequenz für deine Power-Platform-Governance ist simpel und unbequem: Leg jetzt fest, wer für welchen Maker-Agenten bezahlt, wie viel er verbrauchen darf und wer entscheidet, wenn das Budget am 15. des Monats aufgebraucht ist. Wer das nicht vorher klärt, klärt es hinterher im Eskalationsmeeting.

Fakten: Das Wichtigste in 30 Sekunden

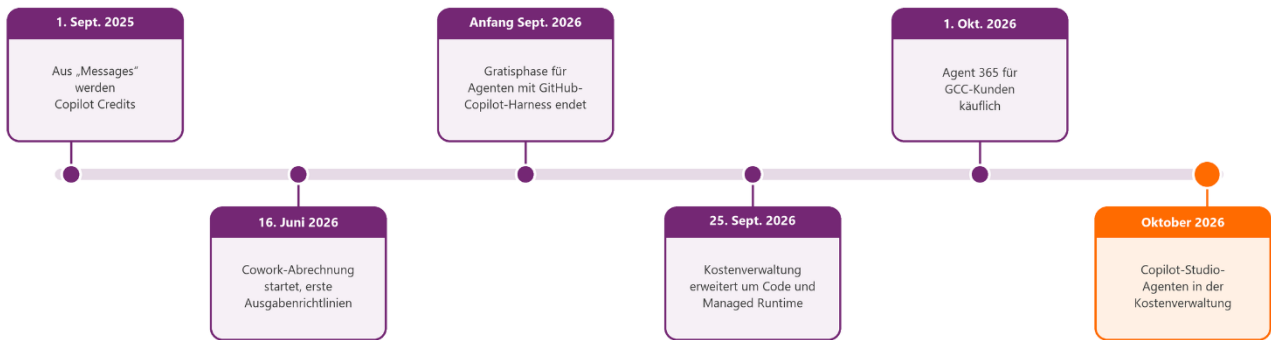
- **Bisher abgedeckt:** Copilot Cowork und Work IQ APIs, verwaltet unter Copilot > Cost Management im Microsoft 365 Admin Center.
- **Neu seit Ende September:** Code und Copilot Managed Runtime.
- **Geplant für Oktober 2026:** Agenten aus Copilot Studio.
- **Neu dazu:** Ausgabenrichtlinien lassen sich per API in großem Stil setzen, Credit-Anfragen von Nutzern laufen in bestehende Genehmigungsworkflows.
- **Währung:** Copilot Credits, im Pay-as-you-go-Modell rund 0,01 US-Dollar pro Credit.

Worum geht es im Detail?

Ein bisschen Hintergrund, damit du verstehst, warum Microsoft hier aufräumt. Seit dem 1. September 2025 heißt die Agenten-Währung nicht mehr „Messages“, sondern **Copilot Credits**. Ein Credit misst grob, wie viel Arbeit ein Agent leistet: Eine klassische Antwort kostet einen Credit, eine generative Antwort zwei, eine Agentenaktion fünf. Werkzeuge und Sprachfunktionen liegen darüber. Mit Cowork kam im Juni 2026 dann die nutzungsbasierte Abrechnung im großen Stil:

Jede Aufgabe kostet, was Modell, Kontextabruf, Werkzeugaufrufe und Laufzeit zusammen verbrauchen. Ein Pauschaltarif sieht anders aus.

Vom Zählwerk zur Budgetbehörde: die Kostenspür der Agenten



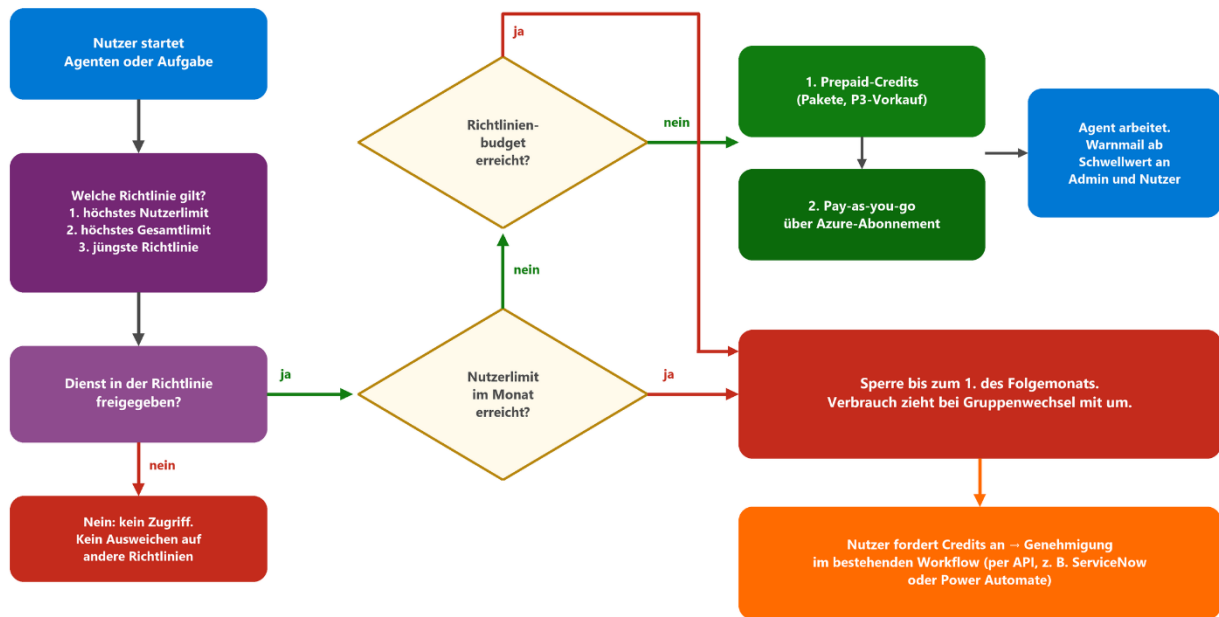
Termine laut Microsoft-Ankündigungen. „Geplant“ heißt bei Microsoft: irgendwann zwischen jetzt und Weihnachten.

Abbildung 1: Wie aus einem Zählwerk eine zentrale Budgetsteuerung wurde

Parallel dazu hat Microsoft im Admin Center das **Cost Management** aufgebaut. Dort legst du Ausgabenrichtlinien an, die drei Dinge regeln: wer (alle Nutzer oder Sicherheitsgruppen), was (welche Agenten und Dienste) und wie viel (unbegrenzt oder begrenztes Monatsbudget, optional mit Obergrenze pro Nutzer). Dazu kommen Warnschwellen für Admins und Nutzer, Modellprofile für Cowork und die Wahl der Abrechnungsmethode. Global- und Abrechnungsadministratoren dürfen pro Richtlinie ein anderes Azure-Abonnement hinterlegen. Das ist nichts anderes als Kostenstellenrechnung für KI, und dein Controlling wird das lieben.

Interessant sind die Details, die erst auffallen, wenn es knallt. Ist ein Nutzer in mehreren Richtlinien, gewinnt die mit dem höchsten Nutzerlimit, bei Gleichstand die mit dem höchsten Gesamtlimit, danach die jüngste. Werte werden nicht kombiniert. Ist das Limit erreicht, ist bis zum Ersten des Folgemonats Schluss, der Nutzer rutscht nicht in eine andere Richtlinie. Wechselt jemand die Gruppe, zieht sein Verbrauch mit um. Der Trick „Ich lass mich mal schnell in die Vertriebsgruppe aufnehmen, die hat noch Budget“ funktioniert also nicht. Schade für die Kreativen.

Was passiert, wenn ein Maker-Agent Credits verbrennt?



Richtlinien reservieren keine Credits, sie deckeln nur. Wer zuerst kommt, verbrennt zuerst.

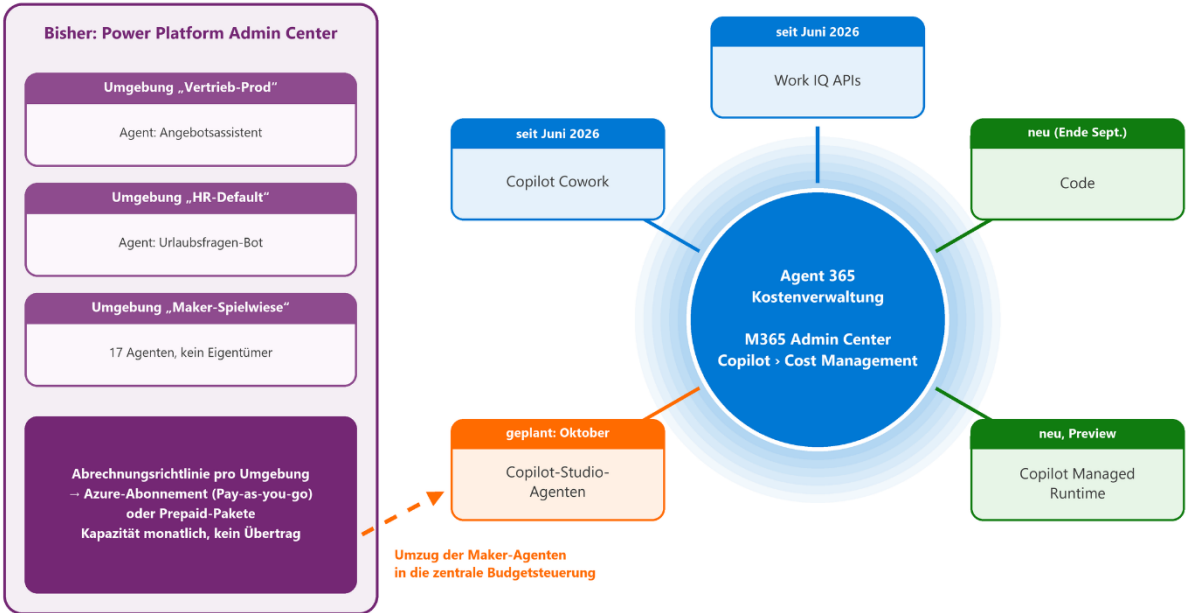
Abbildung 2: Entscheidungslogik einer Ausgabenrichtlinie, vom Agentenstart bis zur Sperre

Und jetzt Copilot Studio. Heute rechnet ein Maker-Agent je nach Umgebung ab: per Prepaid-Paket (etwa 25.000 Credits für rund 200 US-Dollar im Monat), per Vorkaufplan mit einjähriger Bindung oder per Pay-as-you-go über eine Abrechnungsrichtlinie im Power Platform Admin Center. Kapazität gilt monatlich und verfällt, Überziehung kann zur technischen Sperre führen. Nutzer mit Microsoft-365-Copilot-Lizenz verbrauchen in Copilot Chat, Teams und SharePoint für klassische und generative Antworten sowie Graph-Grounding nichts. Sobald ein Agent aber Aktionen ausführt, Flows anstößt oder auf einer Website für Kunden plaudert, tickt der Zähler.

Bleibt die Frage, warum ausgerechnet **Code** und die **Copilot Managed Runtime** jetzt mit in die Kostenverwaltung wandern. Code ist Microsofts Werkzeug, mit dem Nutzer per Copilot kleine Anwendungen und Automatisierungen bauen. Seit Ende September rollt es für Frontier-Kunden aus, die breite Verfügbarkeit soll in den kommenden Wochen folgen. Die Managed Runtime ist die passende Hosting-Umgebung dazu: Sie führt diesen Code innerhalb deiner Microsoft-365-Umgebung aus, unter Kontrolle der IT, und ist derzeit in der Preview. Beides läuft nutzungsbasiert. Übersetzt heißt das: Microsoft hat jedem Fachbereich gerade eine Werkbank mit Taxameter hingestellt. Dass die Budgets dafür an derselben Stelle landen wie Cowork, ist nur konsequent.

Aspekt	Heute: Power Platform Admin Center	Ab Oktober: Agent 365 Kostenverwaltung
Steuerungseinheit	Umgebung	Nutzer, Sicherheitsgruppe, Tenant
Budget	Kapazitätszuweisung, Azure-Budgets	Monatslimit pro Richtlinie und pro Nutzer in Credits
Bei Überschreitung	Technische Durchsetzung bis zur Dienstverweigerung	Sperre bis Monatsersten, Credit-Anfrage mit Genehmigung
Kostenstelle	Azure-Abonnement je Umgebung	Abrechnungsmethode je Richtlinie
Zuständig	Power-Platform-Admin	KI-Admin, Lizenzadmin, Abrechnungsadmin
Automatisierung	PowerShell, Admin-Connectoren	API für Richtlinien, Anbindung an Genehmigungsworkflows

Zwei Kassen, ein Kassensturz: Wohin die Copilot Credits ab Oktober laufen



Stand 11.10.2026. Copilot-Studio-Unterstützung ist angekündigt, nicht ausgerollt. Prüfe dein Message Center, bevor du feierst.

Abbildung 3: Die zwei bisherigen Kostenwelten und ihr Zusammenführen in Agent 365

Wichtig: Angekündigt ist nicht ausgerollt

Microsoft spricht von „geplant für Oktober“. Wie genau Copilot-Studio-Agenten mit den bestehenden Abrechnungsrichtlinien pro Umgebung zusammenspielen, ist noch nicht dokumentiert. Rechne damit, dass beide Welten eine Weile nebeneinander existieren, und plane keine Abschaltung des alten Modells, bevor du die neue Variante im eigenen Tenant gesehen hast.

Was sind Chancen? Was sind Risiken?

Die größte Chance ist banal: **endlich eine Sicht auf alles**. Wer heute wissen will, was KI im Haus kostet, öffnet drei Portale, exportiert vier CSV-Dateien und bastelt sich in Excel eine Wahrheit, die beim nächsten Meeting schon wieder falsch ist. Mit Copilot-Studio-Agenten in der Agent-365-Kostenverwaltung siehst du Verbrauch nach Nutzer, Gruppe, Dienst und Agent an einer Stelle.

Ein Kunde aus dem Maschinenbau hat das schmerzhaft gelernt. Ein Werkstudent baute im Frühjahr einen Agenten, der Ersatzteilanfragen beantwortet, inklusive Aktion zur Lagerabfrage. Der Agent war gut. So gut, dass ein Händler ihn per Skript im Minutentakt befragte. Die Umgebung hing per Pay-as-you-go an einem Azure-Abonnement, das Budget-Alarm hatte niemand eingerichtet. Aufgefallen ist es dem Controlling sieben Wochen später. Der Werkstudent war da schon zurück an der Uni, der Agent lief weiter, unsterblich wie ein Zombie mit Kreditkarte.

Es geht auch anders. Eine Versicherung, die wir begleiten, hat ihren Schadenmelde-Agenten von Anfang an mit eigener Kostenstelle und einem Monatsbudget betrieben. Nach drei Monaten lagen Kosten und eingesparte Sachbearbeiterstunden nebeneinander auf einer Folie, und der Vorstand hat das Budget verdoppelt, statt es zu streichen. Transparenz ist eben nicht nur eine Bremse, sondern auch das beste Argument für mehr Geld. Genau diese Rechnung wird mit der zentralen Kostenverwaltung für jeden Maker-Agenten möglich.

Chancen	Risiken
Eine Konsole für Cowork, Code, Runtime und Copilot Studio	Übergangsphase mit doppelter Logik: Umgebungsrichtlinie und Ausgabenrichtlinie
Budgets pro Abteilung über Gruppen und eigene Azure-Abonnements	„Auto-apply new services“ ist standardmäßig an: Neue Dienste landen still in bestehenden Budgets
Credit-Anfragen laufen in vorhandene Genehmigungsworkflows	Harte Sperre am Limit trifft auch geschäftskritische Agenten
Richtlinien per API: Governance als Code statt Klickstrecke	Richtlinien reservieren nichts: Ein Vielnutzer kann das Gruppenbudget allein leeren
Vergleichbarkeit: Agent-Nutzen gegen Agent-Kosten	Unklare Zuständigkeit zwischen Power-Platform-Team und Microsoft-365-Team

Warnung: Der Schalter, den keiner bemerkt

Neue Ausgabenrichtlinien haben „**Auto-apply new services**“ standardmäßig aktiv. Sobald Microsoft Copilot-Studio-Agenten freischaltet, können diese automatisch in deine bestehenden Cowork-Budgets rutschen. Dann frisst der HR-Urlaubsbot das Budget, das eigentlich für die Vertriebsanalysen gedacht war. Prüfe jede Richtlinie jetzt und schalte die Option dort ab, wo du keine Überraschungen willst.

Das zweite Risiko ist organisatorisch. Die harte Sperre ist ein Segen für das Budget und ein Fluch für den Betrieb. Wenn der Kundenservice-Agent am 20. des Monats verstummt, weil die Gruppe ihr Limit erreicht hat, nützt dir die Kostentreue wenig. Kritische Agenten brauchen deshalb eigene Richtlinien mit eigenem Budget und einen Eskalationspfad, der schneller ist als das Monatsende. Klingt nach Bürokratie? Ist es auch. Aber die Alternative ist ein Bot, der Kunden mit Schweigen bedient. Das kann die Telefonhotline auch ohne KI.

Was müssen wir jetzt schon vorbereiten?

Die Arbeit beginnt nicht im Admin Center, sondern mit einer Inventur. Du musst wissen, welche Copilot-Studio-Agenten in welchen Umgebungen laufen, wer sie gebaut hat, wer sie noch pflegt und welche Abrechnungsrichtlinie heute dahintersteht. In vielen Tenants ist die Antwort auf die dritte Frage: niemand. Diese Agenten sind die ersten Kandidaten für die Stilllegung, bevor sie Kosten mit Budgetadresse produzieren.

Ein besonderes Augenmerk verdient die **Standardumgebung**. Dort landet alles, was Fachbereiche ohne Projektantrag ausprobieren, und dort hängt meist auch keine saubere Abrechnungsrichtlinie. Nutze die Gelegenheit, produktive Agenten in eigene Umgebungen zu verschieben und die Standardumgebung zur persönlichen Spielwiese mit engem Budget zu erklären. Wer dort einen geschäftskritischen Agenten betreibt, hat ihn ohnehin auf Sand gebaut.

Schritt	Was konkret	Warum
1. Agenteninventar	Alle Copilot-Studio-Agenten je Umgebung mit Ersteller, Eigentümer, Kanal und Verbrauch der letzten drei Monate erfassen	Ohne Liste keine Budgetzuordnung
2. Eigentümer benennen	Je Agent einen fachlichen Eigentümer und eine Kostenstelle festlegen, verwaiste Agenten abschalten	Wer zahlt, entscheidet über Nutzen
3. Gruppen bauen	Sicherheitsgruppen je Abteilung und für kritische Agenten anlegen	Richtlinien greifen nur über Gruppen
4. Budgets ableiten	Agent-Nutzungsschätzer und Ist-Verbrauch nutzen, Monatslimit plus Nutzerlimit festlegen	Damit ein Vielnutzer nicht alles leert
5. Auto-apply prüfen	In allen bestehenden Ausgabenrichtlinien bewusst entscheiden	Verhindert stille Budgetverschiebungen
6. Genehmigungsweg	Credit-Anfragen in bestehenden Workflow einbinden, Antwortzeit festlegen	Sperre am Limit braucht schnellen Ausweg
7. Rollen klären	KI-Admin, Lizenzadmin, Abrechnungsadmin und Power-Platform-Admin abstimmen	Zwei Teams, eine Kasse

Tipp: Rechne einmal mit echten Zahlen

Ein Agent mit 2.000 Unterhaltungen im Monat, je zwei generativen Antworten und einer Aktion, verbraucht rund 18.000 Credits, im Pay-as-you-go also etwa 180 US-Dollar. Klingt harmlos. Multipliziere das mit den 40 Agenten aus der Maker-Spielwiese und du hast ein Gesprächsthema für den nächsten Budgetzyklus. Der offizielle Copilot Studio Agent Usage Estimator hilft beim Hochrechnen.

Wichtig: Governance als Code

Weil Microsoft Ausgabenrichtlinien künftig per API setzen lässt, solltest du sie wie jede andere Konfiguration behandeln: versioniert, dokumentiert und reproduzierbar. Wer seine Budgets nur im Portal zusammenklickt, weiß in sechs Monaten nicht mehr, warum der Einkauf ein dreifaches Limit hat.

Und vergiss die Menschen nicht. Deine Maker müssen wissen, dass ihr Agent ab Oktober nicht mehr unsichtbar im Hintergrund Kosten erzeugt, sondern in einem Bericht mit Namen und Betrag auftaucht. Das wirkt erfahrungsgemäß disziplinierender als jede Richtlinie. Kündige es freundlich an, bevor es jemand im Bericht selbst entdeckt. Niemand liest seinen Namen gern zum ersten Mal neben einer vierstelligen Summe.

Häufig gestellte Fragen

Wo finde ich die Kostenverwaltung für Copilot-Agenten im Microsoft 365 Admin Center?

Die Kostenverwaltung liegt im Microsoft 365 Admin Center unter Copilot > Cost Management. Dort legst du auf dem Reiter „Configuration“ Ausgabenrichtlinien an und siehst auf dem Verbrauchsreiter die Copilot Credits nach Nutzer, Gruppe, Dienst und Agent.

Ab wann werden Copilot-Studio-Agenten in der Agent-365-Kostenverwaltung unterstützt?

Microsoft hat die Unterstützung für Agenten aus Copilot Studio für Oktober 2026 angekündigt. Bis die Funktion im eigenen Tenant sichtbar ist, gilt weiterhin die Abrechnung über Abrechnungsrichtlinien und Kapazität im Power Platform Admin Center.

Was passiert, wenn ein Nutzer sein Copilot-Credit-Limit erreicht?

Der Nutzer verliert bis zum Ersten des Folgemonats den Zugriff auf die Agenten und Dienste der betroffenen Ausgabenrichtlinie. Er kann zusätzliche Credits anfordern, wechselt aber nicht automatisch in eine andere Richtlinie.

Wer darf Ausgabenrichtlinien für Copilot Credits anlegen?

Ausgabenrichtlinien, Limits und Warnungen dürfen KI-Administratoren und Lizenzadministratoren verwalten. Die Abrechnungsmethode, also etwa ein bestimmtes Azure-Abonnement, dürfen nur Global- und Abrechnungsadministratoren festlegen oder ändern.

Kosten Copilot-Studio-Agenten extra, wenn Nutzer eine Microsoft-365-Copilot-Lizenz haben?

Für lizenzierte Nutzer sind klassische und generative Antworten sowie Microsoft-Graph-Grounding in Copilot Chat, Teams und SharePoint ohne zusätzliche Credits enthalten. Agentenaktionen, Flows und Agenten auf externen Kanälen verbrauchen dagegen Copilot Credits.