Microsoft

Agent Cost Management

1 Introduction

If you are reading this, you likely face a challenge: your organization is beginning to deploy AI agents, or perhaps you are already witnessing significant usage, and you are uncertain about how to predict or control the associated costs. Consumption-based billing, including prepaid packs and pay-as-you-go meters, offers great flexibility but can also result in unexpected charges, complex budgeting, and governance challenges.

You are not alone. Many IT leaders, administrators, and cost managers encounter similar questions:

- How do we estimate costs before deployment?
- How do we track our expenditures?
- How can we set up alerts or controls to avoid budget overruns?
- How can we regulate who can use how much?

This e-book provides answers. It explains the full lifecycle of agent cost management, from estimation and billing setup to tracking, forecasting, and active cost control. Whether you are just beginning or refining your approach at scale, this e-book offers actionable practices, useful tools, and links to trusted Microsoft Learn content to help you manage agent costs with confidence.

1.1 Problem Statement/Purpose

In this e-book we will showcase the cost reporting and estimation tools needed to manage agents using licenses and/or consumptive billing (via prepaid message packs and pay-as-you-go (PAYG) meters).

As organizations adopt AI-driven agents to enhance productivity and automate workflows, understanding and managing the costs associated with these agents has become critical. The introduction of consumption-based billing models, such as prepaid message packs and PAYG meters, presents both opportunities for flexible adoption and challenges in financial predictability. Without clear strategies for estimation, tracking, forecasting, and management, agent costs can escalate rapidly, potentially leading to budget overruns and governance concerns.

This e-book aims to serve as a practical manual for agent cost management. It not only explains what organizations need to consider and why these considerations matter but also provides actionable guidance on how to implement effective cost management practices. Drawing on lifecycle recommended practices, real-world examples, and Microsoft resources, this e-book empowers IT and business leaders to make informed decisions, optimize costs, and maintain control over their AI investments.

1.2 Scope

This e-book focuses on cost management for agents built and operated across Microsoft platforms, specifically agents created with Copilot Studio, Copilot Studio Agent Builder and pro developer tools (including Teams Toolkit and Azure AI Foundry).

It covers the full agent cost management lifecycle:

- Estimating initial consumption-based costs.
- Selecting and setting up billing models (User License, PAYG, Prepaid).
- Tracking and analyzing actual usage and spending.
- Forecasting future costs based on trends and historical data.

Applying governance controls to manage, optimize, and scale usage responsibly.

Out of scope for this e-book:

- Broader AI governance and data compliance topics (covered in the Governance Whitepaper).
- Technical details of agent development and architecture.
- Non-Microsoft agent platforms.

This e-book complements existing Microsoft Learn documentation and links directly to relevant resources for more information.

1.3 Target Audience

This e-book is intended for roles involved in managing the financial, operational, and governance aspects of agent adoption and usage in an organization:

- **Microsoft 365 Admins** – responsible for managing agent billing, usage tracking, and governance through Microsoft 365 Admin Center and Copilot Control System.
- **Power Platform Admins** – responsible for managing Power Platform based agents, including billing plans, consumption tracking, and cost optimization.
- **IT Leaders and Cost Managers** – seeking to understand and control the financial impact of AI agent adoption and usage.
- **Business Unit Leaders** – needing to align departmental usage and costs with budgets and operational goals.
- **Governance and Compliance Officers** – ensuring cost management aligns with broader governance frameworks and risk management.

This e-book assumes a working knowledge of Microsoft 365, Power Platform, and Azure administration, but will link to introductory resources as appropriate.

1.4 How to use this e-book

This e-book is structured to provide both **conceptual understanding** and **practical instructions** for managing agent costs throughout their lifecycle. It can be read end-to-end for a comprehensive understanding or used as a reference for specific cost management tasks.

We recommend the following approach:

1. **New to agent cost management?**
   Begin with the *Introduction* and *Overview* sections to understand the key concepts, the lifecycle of cost management, and how agents and governance zones fit together.
2. **Planning or budgeting for agents?**
   Use the *Estimating Agent Costs* and *Setting Up Billing Models* sections to guide financial planning and select the right billing model for your needs.
3. **Configuring or adjusting billing?**
   Refer to the *Setting Up Billing Models* sections for step-by-step instructions on enabling and configuring billing models such as User Licenses, Prepaid, and pay-as-you-go in Microsoft 365 Admin Center and Power Platform Admin Center.

4. **Monitoring and optimizing usage?**
   The *Tracking, Analyzing, and Forecasting Agent Costs* section explains how to interpret usage data and adjust billing or governance settings accordingly.
5. **Managing financial risk or governance?**
   The *Managing Agent Costs* section provides techniques for applying controls (limits, policies, alerts) to prevent overages and enforce accountability.

Throughout the e-book, you will find:

- **Lifecycle recommended practices** for each phase of cost management.
- **Links to Microsoft Learn** for detailed configuration guidance.
- **Scenario-based recommendations** for applying governance controls.
- **Notes on coming soon product features** where possible upcoming features will expand your capabilities.

Finally, we recommend that organizations adopting agents at scale consider this e-book as a living guide. As Microsoft introduces new tools (such as the Copilot Control System) and updates billing features, revisit this e-book regularly to refine your cost management strategies.

2 Overview

Agents are intelligent software entities designed to perform specific tasks, streamline workflows, and enhance user productivity. These agents leverage predefined algorithms, rules, or artificial intelligence models to interact with users and systems in meaningful ways. They can automate processes, provide insights, and facilitate decision-making, making them indispensable tools in modern organizational contexts.

Agents can be built using:

- **Copilot Studio Agent Builder** – Use Copilot Studio Agent Builder tool directly within Microsoft 365 Copilot Chat in Teams to autonomously handle tasks, answer questions, and interact with enterprise data to support users in real time.
- **Copilot Studio** – Use triggers, advanced logic, and connections to other Microsoft services or third-party platforms to create agents with full Copilot Studio.
- **Pro Developer Tools** – Use tools and services like Teams Toolkit, Copilot Studio native integration with Visual Studio, and Azure AI Foundry to build fully-customized agents with the model and orchestration engine of your choice.

Zones represent distinct levels or stages of governance maturity within the Microsoft 365 ecosystem. Each zone is designed to address specific security, management, and operational needs while ensuring a scalable and adaptable framework for agents. By conceptualizing governance through zones, organizations gain a clearer pathway to evolve their strategies, starting from foundational controls to advanced, fully integrated solutions.

The zones serve as a structured roadmap for IT practitioners and decision-makers to implement and manage agents effectively, aligning security and operational measures with their organization's maturity level. Each zone emphasizes key elements such as access management, compliance enforcement, resource optimization, and system integration. Organizations can progress through these zones by enhancing governance practices, expanding security measures, and refining management strategies. This tiered approach ensures adaptability to meet growing organizational demands while maintaining a robust governance structure.

The zones are defined as being:

- **Zone 1**, the Personal Productivity Zone which serves as the entry point in the structured governance framework, providing organizations with the tools and guidelines necessary to manage their agents effectively and securely within the Microsoft 365 ecosystem.
- **Zone 2**, the Collaboration Zone builds upon Zone 1 and represents a step forward with governance maturity, empowering organizations to maintain operational excellence while adapting to the complexities of modern IT environments.
- **Zone 3**, the Enterprise Managed Zone builds upon Zone 2 and represents an advanced governance framework stage that focuses on optimizing IT operations through enhanced security, sophisticated management protocols, and predictive analytics. It integrates technology-driven solutions such as advanced threat detection, multi-factor authentication, and continuous monitoring while supporting complex agent scenarios with scalable management controls and actionable reporting for strategic decision-making.

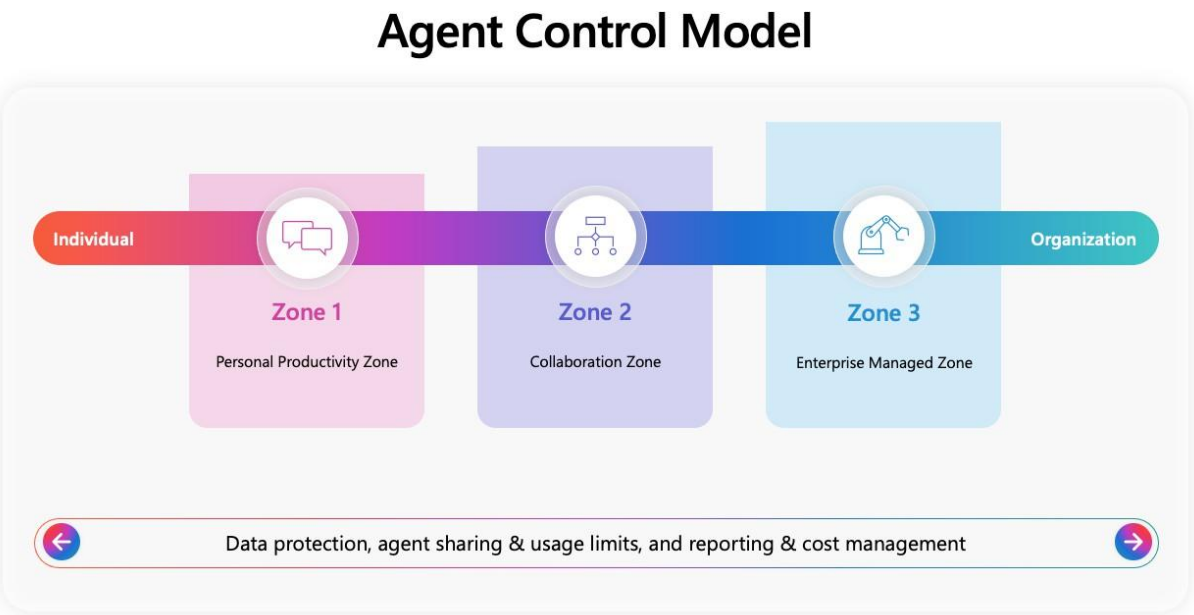

Figure 1 Agent Zones

The lifecycle of cost management encompasses several critical stages that organizations undertake to ensure effective financial oversight within their systems. These stages are

designed to streamline expenses, enhance forecasting accuracy, and align cost structures with operational goals.

The first stage involves **estimating costs**, a process where organizations predict expenses based on anticipated resource usage and operational needs. This step requires the integration of historical data, market trends, and consumption estimates to create a reliable baseline for budgeting.

Next, organizations proceed to **set up a Billing Model**. This includes selecting the appropriate payment model, whether it's User License, Pay-As-You-Go (PAYG), or Prepaid. A Billing Model acts as the financial foundation, enabling clear tracking and categorization of expenditures.

Once the Billing Model is established, the focus shifts to **analyzing and forecasting costs**. This stage leverages advanced analytics tools to monitor consumption and predict future spending patterns. By identifying areas of inefficiency and opportunities for optimization, organizations can better manage their budgets and adapt to changing operational demands.

Finally, **managing agent costs** forms the operational phase. This involves continuous oversight of expenditures related to software agents, ensuring alignment with organizational objectives while maintaining cost-effectiveness. Through this lifecycle, organizations can achieve a balanced approach to cost management, promoting financial sustainability alongside operational excellence.

Learn more about Copilot Studio agent builder at Use Copilot Studio Agent Builder to Build Agents | Microsoft Learn.

Learn more about Microsoft 365 Copilot Chat at Overview of Microsoft 365 Copilot Chat | Microsoft Learn.

Learn more about Copilot Studio at Microsoft Copilot Studio guidance documentation - Microsoft Copilot Studio | Microsoft Learn.

3 Estimating Agent Costs

Before adopting AI agents at scale, it is essential to estimate their potential costs. The flexible and usage-based billing models now available – Prepaid, Pay-As-You-Go (PAYG), and User License – offer agility but also introduce financial unpredictability if not well planned.

You should estimate agent costs during the planning phase for any new custom engine agent, when preparing financial forecasts for AI programs, or as part of pre-sales discussions for AI adoption at either a departmental or enterprise level. Estimation is particularly important before enabling agents in production environments or rolling them out across a broad user base.

By accurately estimating agent costs during the planning phase, organizations can align budgets, manage stakeholder expectations, and avoid surprises after deployment.

Learn more about agents at Custom engine agents for Microsoft 365.

3.1 Why Estimation Matters

Without proactive estimation, organizations may encounter budget overruns, governance gaps, and stakeholder dissatisfaction. Effective estimation:

- **Supports business case development** – helps justify AI investments to leadership and finance teams.
- **Aligns budgeting with expected usage** – prevents unplanned budget shortfalls or approvals delays.
- **Guides choice of billing model** – helps determine if PAYG or Prepaid packs offer better value at the current scale.
- **Informs governance planning** – enables proactive decisions on agent scope, limits, and guardrails.

3.2 How to Measure Agent Costs

To assist with estimation, Microsoft has released the **Copilot Studio Agent Consumption Estimator** tool at https://aka.ms/CopilotStudioEstimator.

This tool helps customers and partners **estimate and project their monthly consumption of a custom engine agent**.

It offers:

- Forecasted monthly message consumption based on typical usage scenarios.
- Cost comparisons for Prepaid and PAYG billing models.
- Visibility into what messages may be zero-rated under Microsoft 365 Copilot licenses.

For enterprise-wide forecasting, the Copilot Studio Agent Consumption Estimator is designed to complement (not replace) tools such as Microsoft's Value Estimation Tool (VET), which evaluates Copilot and Copilot Chat investments across an organization.

3.3 Estimating Other Agent Costs

While the Copilot Studio Agent Consumption Estimator is an excellent starting point for custom engine agents, broader agent cost estimation should also include:

Agent Type	Estimation Approach
M365 Copilot Chat	Typically covered under Microsoft 365 licensing (no additional consumption cost)
SharePoint-based agents	Covered under Microsoft 365 licensing (no additional consumption cost)
Dynamics 365-based agents	Estimate based on Dynamics 365 licensing and applicable add-ons
Azure AI-based agents	Estimate using Azure Consumption API or Azure Cost Estimator tools

Table 1 Estimation approaches for agent types

3.4 Recommended Practices

Estimating agent costs is not a one-time exercise — it should be an ongoing part of your AI governance and financial planning process. Based on learnings from early

customer deployments and Microsoft internal experience, the following recommended practices can help ensure your cost estimates remain accurate and useful as adoption scales:

- Begin with **conservative estimates**, as adoption and usage often ramp quickly once agents demonstrate value.
- Model for **peak usage**, not just averages — this helps prevent unexpected costs during periods of high demand.
- Take into account **seasonality** or predictable business cycles that may cause spikes (e.g., fiscal year-end, customer service peaks).
- Review and update your estimates on a **quarterly basis**, especially as new agent features (such as multi-agent orchestration or code interpreter) become available and may impact consumption patterns.

4 Setting Up Billing Models

Agents use "messages" as a common unit of usage across its capabilities. To meet varying needs, it offers three licensing options to give you flexibility based on your usage.

- **Pre-paid** – for more predictable usage, there's the Message Pack option. It's US$200 per tenant per month for 25,000 messages. You can buy it directly in the Microsoft 365 admin center. The message pack is applied at the tenant level, and you can buy multiple packs as needed.
- **Pay-as-you-go (PAYG)** – great for unpredictable or low-volume usage. You pay US$0.01 per message, and it's billed through Azure. No upfront purchase is needed.
- **User License** – Copilot Studio is also included in the Microsoft 365 Copilot license, which is US$30 per user per month. This gives each licensed user full access to build agents. It's available via your Microsoft licensing provider or in the Microsoft 365 admin center. Agents built in Copilot Studio for Microsoft Teams, SharePoint, and Microsoft 365 Copilot are included at no extra charge.

Each plan offers flexibility depending on how you want to scale and control your Copilot Studio usage.

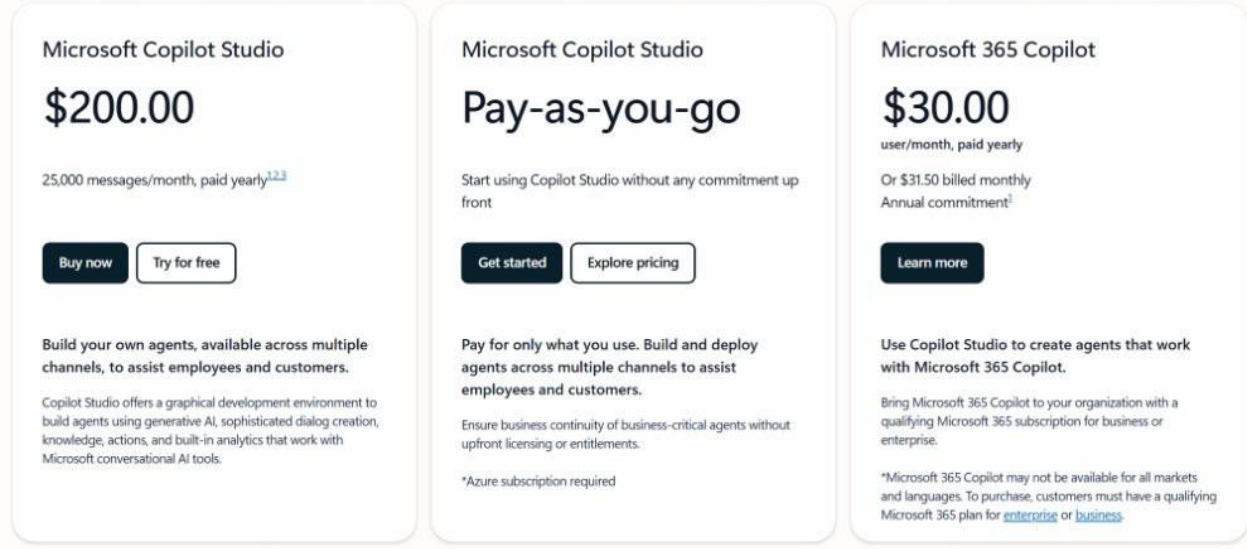

Figure 2 Billing model options

Note that **Copilot Chat is included at no additional cost** with Microsoft 365. It's available natively and no additional setup is required. For users without Microsoft 365 licensing, activate Pay-as-you-go (PAYG) to ensure non-licensed users can easily access Copilot Chat.

To scale Copilot Chat to additional users:

- **Pay-as-you-go (PAYG)**:
  Activate PAYG to access powerful agents that address real business needs—without licensing every user. It's flexible, scalable, and ideal for teams exploring innovation.
  - **Cost**: $0.01USD per message
  - **Billing**: Through Azure, with no upfront purchase, just link an Azure subscription.
- **Copilot Licenses**:
  Assign licenses to users for the full suite of Copilot capabilities across Microsoft 365 apps—Word, Excel, PowerPoint, Outlook, and Teams.
  - **Cost**: $30USD per user/month
  - **Benefits**: Richest, most integrated experience with full access to build and manage Copilot and agents
  - **Availability**: Via your Microsoft licensing provider or the Microsoft 365 admin center

Learn more about Copilot chat at Manage Microsoft 365 Copilot Chat | Microsoft Learn.

4.1 Adoption Strategy

The Microsoft licensing model is designed to allow organizations to start exploring AI capability without incurring significant cost or risk. As the business benefit is established organizations can quickly scale up to other license models which offer better affordability at scale (while ensuring other governance factors have been addressed).

This provides a means for an organization to adopt agents in three basic steps that ensure compliance with a variety of regulatory frameworks while managing costs:

- Start with PAYG — this helps teams learn, validate governance needs, and demonstrate business value.

- Migrate to a combination of Prepaid and PAYG for scalable, affordable consumption as usage ramps.
- Move to User License for well-established use cases where per-user billing provides more predictable costs.

To change your Billing Model, refer to *Change between Billing Models*.

4.2 Setting Up User Licenses

Organizations can assign Microsoft 365 Copilot licenses directly to users through the Microsoft 365 Admin Center. These licenses provide full rights to build and manage Copilot Studio agents across Teams, SharePoint, and M365 Copilot Chat — with zero-rated message consumption for declarative agents.

To assign licenses:

- Navigate to the **Microsoft 365 Admin Center**
- Go to **Billing** → **Licenses**
- Select **Microsoft 365 Copilot**
- Assign licenses to individual users or security groups

For detailed steps, see: Set up Microsoft 365 Copilot and assign licenses | Microsoft Learn.

4.3 Setting Up Prepaid

If your organization purchases **Prepaid Copilot Studio message packs**, you can distribute that capacity across environments to support specific departments or business units.

In the **Power Platform Admin Center** (PPAC), admins can:

- **Allocate message capacity** – to production or sandbox environments (e.g., Sales, Marketing, Support).
- **Isolate allocations** – environments draw only from their assigned pool.
- **Prevent cross-environment usage** – e.g., Marketing can't "borrow" from Sales unless reconfigured by an admin.

This supports clear cost accountability, departmental billing, and chargeback models. This is especially useful in organizations with **chargeback models** or **cost accountability**, where each team is responsible for their own usage.

Learn more about setting up Prepaid at Manage Copilot Studio messages and capacity - Power Platform | Microsoft Learn.

4.3.1 Coming Soon

Message Capacity Packs will soon be available in the Microsoft 365 Admin Center (MAC) for Microsoft 365 Copilot Chat. This new capability will allow Global admins to apply pre-purchased message capacity—available as monthly or annual subscriptions—before incurring pay-as-you-go (PAYG) charges. Admins will be able to allocate these packs by department or group through billing policies, enabling more granular control over usage and spend management.

Learn more about this coming soon feature at Microsoft 365 Roadmap | Microsoft 365.

4.4 Setting up PAYG

Both the **Microsoft 365 Admin Center (MAC)** and **Power Platform Admin Center (PPAC)** support configuration of pay-as-you-go (PAYG) billing for agent consumption.

Before you begin, ensure that the following prerequisites have been satisfied:

- You have an active Azure subscription in the same tenant as your Microsoft 365 (M365) environment.
- A resource group exists in that subscription.
- You hold the necessary roles:
    - Global Administrator or SharePoint Administrator in M365.
    - Owner or Contributor in Azure for the subscription and resource group.

4.4.1 Setup SharePoint agents and Microsoft 365 Copilot Chat Services PAYG

To configure Pay-as-you-go (PAYG) billing for agents across Microsoft 365 surfaces—including Copilot Chat, SharePoint, and Teams—regardless of whether the agent was created using Copilot Studio, Agent Builder, or Teams Toolkit, follow these high-level steps:

1. In the Microsoft 365 Admin Center, create a billing policy linked to an Azure subscription and resource group.
2. Configure which departments or user groups are included under the billing policy.
3. Enable PAYG billing for selected services (e.g., Microsoft 365 Copilot Chat or SharePoint agents).

For detailed steps learn more at Set up Microsoft 365 Copilot pay-as-you-go for IT admins | Microsoft Learn.

4.4.2 Setup Copilot Studio Agent PAYG

To configure PAYG for **Copilot Studio agents** the high-level steps are:

1. Create a **billing plan** in PPAC — this links an environment to an Azure subscription.
2. Link the subscription and resource group to the billing plan.
3. Associate the billing plan with one or more production or sandbox environments.

For detailed steps learn more at Set up a pay-as-you-go plan - Power Platform | Microsoft Learn.

4.5 Setting Up User Licenses

Organizations can assign Microsoft 365 Copilot licenses directly to users through the Microsoft 365 Admin Center. These licenses provide full rights to build and manage Copilot Studio agents across Teams, SharePoint, and M365 Copilot Chat — with zero-rated message consumption for declarative agents.

To assign licenses:

- Navigate to the **Microsoft 365 Admin Center**.
- Go to **Billing** | **Licenses**.
- Select **Microsoft 365 Copilot**.
- Assign licenses to individual users or security groups.

For detailed steps, learn more at Set up Microsoft 365 Copilot and assign licenses | Microsoft Learn.

5 Tracking, Analyzing, and Forecasting Agent Costs

Once agents are in use across your organization, ongoing tracking and analysis of actual costs is critical to maintaining control over AI spending and informing future budgeting decisions. This phase of the cost management lifecycle ensures that costs remain aligned with business value and governance goals.

In this section, we explore how to use Microsoft's tools — including Microsoft 365 Admin Center (MAC), Power Platform Admin Center (PPAC), and Azure Cost Management — to track real-world agent consumption, analyze trends, forecast future spending, and identify opportunities for optimization.

5.1 Microsoft 365 Admin Center Agent Costs

The Microsoft 365 Admin Center (MAC), through the **Copilot Control System** (CCS), provides a growing set of tools for tracking, analyzing, and forecasting agent-related costs across Microsoft 365 Copilot Chat. These capabilities can help admins monitor usage patterns, identify high-cost agents or users, and manage consumption proactively.

In addition, admins can track and manage agents through the **CCS Agents** experience in MAC. A dedicated **"Shared Agents" tab** provides a centralized inventory of all agents in the tenant, including those created by end users that haven't been published to the IT Catalog. This tab allows admins to:

- View metadata about each agent (e.g., capabilities, data sources, publishing status).
- Search for agents by name or function.
- Identify agents deployed via public store, direct upload, or internal development.

Learn more about the Copilot Control System at Copilot Control System – Microsoft Adoption.

Learn more about managing agents in the M365 Admin Center at Manage agents for Microsoft 365 Copilot in the Microsoft 365 admin center - Microsoft 365 admin | Microsoft Learn.

Learn more about costs and billing for M365 Copilot PAYG at View costs for Microsoft 365 Copilot pay-as-you-go | Microsoft Learn.

5.1.1 Coming Soon

The Message consumption report now available in Preview helps you manage metered consumption costs for Microsoft 365 Copilot Chat. This report gives you visibility into billed messages associated with your Microsoft 365 Copilot pay-as-you-go billing policies and includes key metrics such as:

- Total messages consumed
- Cumulative and daily time series
- Messages consumed per user, per agent, and per agent-user pair

This report will have near-real-time alerts, for high consumption users who have exceeded a certain number of billed messages over the last 30 days

For example, if there are users that have exceeded 2000 billed messages in the past 30 days, the admin will see an alert card in the M365 Admin Center with a link to the user details section in the report.

Learn more about the message consumption report at Microsoft 365 reports in the admin center - Message consumption - Microsoft 365 admin | Microsoft Learn.

5.2 Power Platform Admin Center Agents Costs

The Power Platform Admin Center (PPAC) is an essential resource for managing and forecasting agent-related costs in Microsoft environments. This tool empowers administrators with robust tracking and analytical capabilities to effectively oversee message consumption and associated expenses. By integrating features like real-time alerts for high-consumption users and detailed consumption reports, PPAC facilitates proactive cost management. Admins can leverage this resource to optimize usage, ensure cost efficiency, and predict future expenditures based on consumption trends.

Learn more about managing Copilot Studio messages and capacity at Manage Copilot Studio messages and capacity - Power Platform | Microsoft Learn.

5.2.1 Cost Tracking

Licensing Hub in PPAC provides detailed visibility into agent-related message usage across environments. Key tracking features include:

- **Tenant-Level Monitoring** – Admins can track message consumption by environment, product, agent, and feature. This includes daily and cumulative usage trends.
- **Billing Policy Views** – Near real-time visibility into metered message consumption is available, with planned enhancements to support consumption per billing policy.
- **Capacity Summary Dashboards** – These show prepaid and pay-as-you-go (PAYG) message units, with breakdowns by environment and usage type.

5.2.2 Cost Analysis

PPAC supports deep analysis of agent costs through:

- **Consumption Drilldowns** – Admins can drill down into usage by agent, user, or agent-user pair to understand high-consumption patterns.
- **Azure Cost Management Integration** – For PAYG setups, Azure provides billing details by meter and resource, enabling granular cost attribution.

- **Custom Reporting** – Downloadable reports from PPAC offer breakdowns by environment, app/agent, and user, supporting internal chargeback models.

5.2.3 Cost Forecasting

Forecasting capabilities are built on historical usage data:

- **Historical Trends:** PPAC provides up to 12 months of monthly data and three months of daily data to support budgeting and licensing planning.
- **Consumption Estimators:** Tools like the Copilot Studio agent consumption estimator help predict future usage based on current patterns.

5.2.4 Coming Soon

Managing agent costs is an ongoing process that requires both proactive governance and responsive controls. As agent usage grows, so does the need for **granular visibility and control** — not just at the environment level, but down to individual agents and users.

Until now, **capacity allocation** (for prepaid packs) and **Pay-As-You-Go (PAYG) billing** have been managed at the environment level. However, as more organizations deploy multiple agents within shared environments, demand has grown for more precise cost controls at the **agent level**.

Coming soon to the PPAC:

- Admins will be able to **define monthly consumption limits** for each Copilot Studio agent — whether the environment uses Prepaid or PAYG.
- A new **Licensing Hub** will provide a centralized view of all agents across the tenant, showing:
    - Configured message limits (if set)
    - Month-to-date usage
    - Associated environments
    - Current usage status (within limit, nearing limit, or over limit)
- Admins will be able to:

- Search for specific agents
- Set usage caps directly
- Enforce **monthly limits** — in Prepaid environments, within the allocated pool; in PAYG environments, with flexible thresholds and overages billed accordingly
- Turn off agents from this interface
- Configure **guardrails** — such as notifications or auto-disabling agents once they reach 100% of their assigned limit

These new capabilities will significantly enhance cost management — helping organizations scale AI adoption while maintaining financial control and avoiding unexpected charges.

5.3 PAYG in Azure

The Azure portal provides robust features for tracking, analyzing, and forecasting PAYG agent costs, enabling organizations to maintain financial control and scale their operations efficiently. These capabilities are designed to enhance cost management, offering both granular insights and proactive tools.

In summary, Azure's portal equips organizations with comprehensive tools for tracking, analyzing, and forecasting agent costs where billing model is chosen as PAYG. By utilizing these capabilities, administrators can enhance cost management, reduce risks of overspending, and optimize operational efficiency in a scalable manner.

5.3.1 Cost Tracking

The Azure portal allows administrators to search for specific agents and directly monitor their consumption patterns. Through the integration of tools like the Azure Consumption API, users can pull detailed usage and billing data, providing real-time visibility into costs. This enables the creation of custom dashboards or portals via the Azure Graph API, allowing organizations to surface real-time metrics and track individual agent costs effectively.

Learn more about Azure cost tracking at Track costs across business units, environments, or projects - Cloud Adoption Framework | Microsoft Learn.

5.3.2 Cost Analysis

Azure's analytical features include the ability to set usage caps, enforce monthly limits, and configure guardrails for financial control. For instance, administrators can establish notifications or auto-disable agents once they reach 100% of their assigned limits. These measures ensure that organizations can stay within budget while having the flexibility to adjust thresholds in PAYG environments. Additionally, integration with tools like Power Automate or Logic Apps facilitates the creation of automated alerts and governance reports, offering deeper insights into cost trends and usage patterns.

Learn more about Azure cost analysis at Get started with Cost Management reporting - Azure - Microsoft Cost Management | Microsoft Learn.

5.3.3 Cost Forecasting

Forecasting is a key feature in the Azure portal, allowing organizations to anticipate costs and avoid unexpected charges. By leveraging real-time data provided by Azure's APIs, administrators can model future usage scenarios and predict expenses. These insights empower organizations to make informed decisions about scaling their AI adoption while maintaining financial accountability.

Learn more about Azure cost forecasting at Common cost analysis uses in Cost Management - Microsoft Cost Management | Microsoft Learn.

5.3.4 Azure APIs

Azure provides APIs for obtaining real-time cost insights:

- **Azure Consumption API** – This API allows you to retrieve detailed usage and billing data. You can access information about services used, quantities consumed, and the associated costs, enabling thorough analysis of your Azure expenditures. It is particularly useful for tracking spending patterns over time, identifying areas where costs can be reduced, and ensuring the efficient use of resources.

- **Azure Graph API** – Utilize this API to create custom portals or dashboards that display real-time metrics tailored to your specific needs. With the Azure Graph API, you can visualize various performance indicators and operational data, making it easier to monitor and manage your cloud infrastructure. By integrating real-time data into these dashboards, you gain a dynamic view of resource utilization and efficiency.
- **Integration** – Integrate these APIs with Logic Apps or Power Automate to construct automated alerts, governance reports, or workflows. This integration allows for proactive management by automating responses to certain conditions, such as budget thresholds being met. Automated governance reports can help maintain compliance with internal policies, while workflows can streamline processes like provisioning and incident response.

These APIs are recommended for those requiring comprehensive breakdowns, real-time dashboards, or integration with corporate BI tools. They provide robust solutions for maintaining visibility and control over Azure spending and performance, thereby enabling better financial management and operational efficiency.

Learn more about Azure Consumption APIs at Azure Consumption REST APIs | Microsoft Learn.

6 Managing Agent Costs

In the sections that follow, we outline the full range of options currently available — in the Microsoft 365 Admin Center (**MAC)**, Power Platform Admin Center (**PPAC)**, and **Azure** — to help you manage agent costs effectively today. We also highlight where new features (like per-agent limits) will enhance these controls in the near future.

6.1 Manage who can create agents

Managing who can create Copilot agents is critical to prevent agent proliferation, manage security risks, and control usage-based costs.

For example, you may identify that an unexpected number of agents are being created, potentially leading to resource management challenges. If this happens then the recommended action is to manage who can create agents, thereby controlling the number and ensuring proper authorization.

Another example is that agents are being created by individuals who are not anticipated or authorized to do so. To resolve this, it is advised to manage the permissions and roles of users, specifically focusing on who can create agents.

To manage who can create agents, you can perform the following:

- Manage Copilot Studio Licenses
- Manage M365 Copilot Chat agent creation

6.1.1 Manage Copilot Studio Licenses

The following types of users can create Copilot Studio agents:

- Users with a Copilot Studio license, either through prepay message packs or a PAYG billing model.
- Users with a Copilot Studio author role.
- Users with a Microsoft 365 Copilot license.

It is recommended that a dedicated Microsoft Intra security group (e.g., “Copilot Studio Makers”) be used to assign licenses and manage access centrally. Only users in this group should be granted access to environments where agents can be developed and

deployed. This can be done in the Power Platform Admin Center, by assigning the security group to the Copilot Studio author setting.

Learn more about licenses and assigning licenses at Copilot Studio licensing and Assign Licenses.

6.1.2 Additional Controls

To further restrict agent creation and publishing:

- Use environment roles (Maker, Admin, etc.) to limit access to Copilot Studio functionality.
- Enable publishing approvals via Pipelines or solution governance to control how agents are released to business units or regions.

Learn more about Power Platform environments at Work with Power Platform environments - Microsoft Copilot Studio | Microsoft Learn.

Learn more about publishing agents at Key concepts - Publish and deploy your agent - Microsoft Copilot Studio | Microsoft Learn.

6.2 Manage who can use agents

Managing who can use agents is an equally important aspect in managing costs.

For example, you may identify that agents are being accessed or utilized by users who are not authorized or intended to use them. As such you should manage user access and define who is eligible to use the agents.

There are three options to enable agents:

- All users
- No users
- Specific users and groups

By limiting agent usage to specific users initially, administrators have greater visibility over message consumption and enables staged rollout of agents by security groups. This also allows adoption to be managed by business unit and region.

Learn more about managing agents at Manage agents for Microsoft 365 Copilot in the Microsoft 365 admin center - Microsoft 365 admin | Microsoft Learn.

6.3 Block an agent

You may need to block agents that fail to comply with organization policies or agents that are consuming resources at unexpectedly high levels leading to inefficiencies or overages. To block an agent, this can be done through the M365 Admin Center. Although care should be taken to ensure that the agent is not performing any critical business functions before blocking.

Note that blocking an agent will prevent users from being able to install it and it will be removed from any user who has already installed it. Blocking an agent will not delete it and it can be reactivated if needed.

Learn more about blocking agents at Block or unblock an agent.

6.4 Manage granular billing

Granular billing (also known as departmental billing) ensures internal cost transparency and accountability by distributing expenses across specific user groups or departments. It allows organizations to track usage and costs per group through Microsoft Intra security groups or individual user assignments, enabling strategic insights into consumption patterns and supporting departmental chargebacks. This structure helps in aligning costs with budgets, fostering better financial management and accountability.

To enable granular billing, Copilot Studio supports billing policies aligned with user groups or departments. As such, admins can create billing policies scoped to specific Microsoft Entra security groups or individual users. Each policy is then linked to an Azure subscription, enabling usage and costs to be tracked per group.

Learn more about creating and managing billing polices at Create and manage billing policies.

6.5 Manage budget and spending limits

Budget and spending limits are essential for maintaining financial control and avoiding unexpected costs. They allow organizations to define guardrails on usage through features such as hard limits, which suspend services when budgets are exceeded, and soft limits, which trigger notifications at custom thresholds. These measures are particularly useful for large organizations with distributed cost centers, ensuring transparency, accountability, and strategic oversight in managing expenses across departments or user groups.

Copilot Studio offers budget management within billing policies to help organizations set guardrails on usage:

- Define hard limits: Suspend services when the budget is exceeded.
- Define soft limits: Trigger email notifications at custom thresholds (e.g., 70%, 90%).

This feature allows proactive control over spending to avoid surprise costs. It is particularly useful for larger organizations with distributed cost centers.

Learn more about spending limits at Configure spending limits for billing policies.

6.5.1 Coming Soon

In the M365 Admin Center Admins will soon be able to define and enforce budget limits for Copilot PAYG services directly within the Microsoft 365 Admin Center (MAC).

Key capabilities will include:

- **Billing Policy Budgets**: Set a dollar limit per billing policy. These policies are linked to Azure subscriptions and can be scoped to specific departments or user groups.
- **Reset Frequency**: Budgets can be configured to be reset on a monthly, quarterly, or yearly.
- **Email Alerts**: Notifications can be triggered at customizable thresholds (e.g., 80%, 90%, 100%) and sent to mail-enabled security groups.

Learn more about this coming soon feature at: Microsoft 365 Roadmap | Microsoft 365.

6.6 Handle Prepaid Overage

When a prepaid message pack runs out in an environment, Makers may encounter publishing errors and agents may fail to respond due to capacity exhaustion.

To cater for this scenario, it is recommended that the following is configured:

1. Enable draw down from tenant-wide prepaid capacity to restore availability.
2. Enable fallback to pay-as-you-go to allow continued usage without interruption.
    a. This needs to be configured in advance

It is also recommended that Administrators:

- Monitor consumption trends[1].
- Proactively configure PAYG for critical environments.
- Reallocate unused capacity from underutilized environments to those in overage.

Learn more about capacity and overage at Manage Copilot Studio capacity and overage.

6.7 Configure per user limits

Per-user message limits are necessary to maintain budget fairness and prevent abuse within a system. These limits allow administrators to set daily or monthly quotas for individual users, ensuring equitable resource allocation. They can be tailored for specific security groups, such as power users or premium departments, and provide detailed visibility into consumption through tools like the Power Platform Admin Center. By implementing per-user limits, organizations can effectively balance access with cost management while preventing overuse or misuse of resources.

To maintain budget fairness and prevent abuse, per-user message limits can be configured within billing policies:

- Set **daily or monthly quotas per user**.
- Apply to select security groups, such as power users or premium departments.

Learn more about setting message limits at Set message limits per user.

[1] Consumption per billing policy is coming soon at https://www.microsoft.com/microsoft-365/roadmap?filters=&searchterms=496139.

6.8 Configure Agent Level Message Limits

Agent-level message limits, also known as "per agent controls," are crucial for cost containment and efficient resource management. They allow organizations to:

- Assign maximum message thresholds to each agent, preventing overuse.
- Monitor usage at the individual agent level for precise control.
- Set automatic triggers to send alerts or shut down agents nearing their limits.

These measures are especially beneficial for high-volume agents, such as those in customer service, and for limiting internal pilot agents during testing phases. By implementing these controls, organizations can manage budgets effectively while ensuring that resources are allocated appropriately.

Learn more about message limits at Set message limits for specific agents.

6.9 Change between Billing Models

Organizations may need to reassess and adapt their billing strategy within Copilot Studio as their usage patterns shift over time, their team size evolves, or their budgetary constraints change. This flexibility allows businesses to select the most suitable billing option that aligns with their operational needs, whether it involves accommodating growth, optimizing cost efficiency, or responding to fluctuating resource demands. By carefully evaluating these factors, organizations can ensure that they maximize value while maintaining effective control over expenditures.

The following sections explain how you change between the billing models including:

- Prepaid
- Pay-as-you-go (PAYG)
- User Licenses

6.9.1 Prepaid to PAYG

If prepaid capacity is underutilized and more flexibility is needed, consider switching to PAYG:

- Reduce or discontinue future prepaid purchases

- Enable PAYG billing via Azure for environments with unpredictable usage
- Monitor with alerts to avoid runaway costs

Learn more about PAYG at Configure PAYG billing plans.

6.9.2 Prepaid to User License

For small teams or departments with limited, steady use, user licenses offer a simplified alternative to bulk prepaid management:

- Reclaim unused prepaid capacity
- Assign fixed user licenses to known agents or makers

Please note that User Licenses cannot be assigned to External users.

Learn more about license types at Compare license types.

6.9.3 PAYG to Prepaid

If central IT needs greater oversight and budgeting, converting from PAYG to prepaid message packs ensures consistent spend:

- Analyze historical consumption to estimate message pack needs
- Purchase prepaid capacity
- Disable PAYG if no longer needed

Learn more about Prepaid add-ons at Manage Copilot Studio messages and capacity - Power Platform | Microsoft Learn.

6.9.4 PAYG to User License

When user activity is steady and fits under fixed license thresholds.
When usage stabilizes across specific users and consistently stays within thresholds, switching to **per-user licenses** may reduce cost volatility:

- Identify users with predictable patterns via usage reporting
- Assign user licenses and remove PAYG tracking from the environment

Please note that User Licenses cannot be assigned to External users.

Learn more about licenses at Assign per-user licenses.

6.9.5 User License to Prepaid

If you're planning to scale up and want budget predictability, convert from individual user licenses to **prepaid capacity**:

- Purchase prepaid message packs through Microsoft 365 admin center or your licensing provider
- Allocate message volume to key environments
- Track consumption via PPAC

Learn more about Prepaid capacity at Buy and manage prepaid capacity.

6.9.6 User License to PAYG

If user activity becomes variable, or if only a few users need sporadic access to agents, converting from user licenses to **PAYG** offers flexibility.

- Deactivate unused user licenses
- Enable PAYG billing for the relevant environments
- This ensures you're only billed for what you use

Learn more about enabling PAYG at Enable PAYG in Power Platform.

Learn more about setting up PAYG at Set up Microsoft 365 Copilot pay-as-you-go for IT admins | Microsoft Learn.