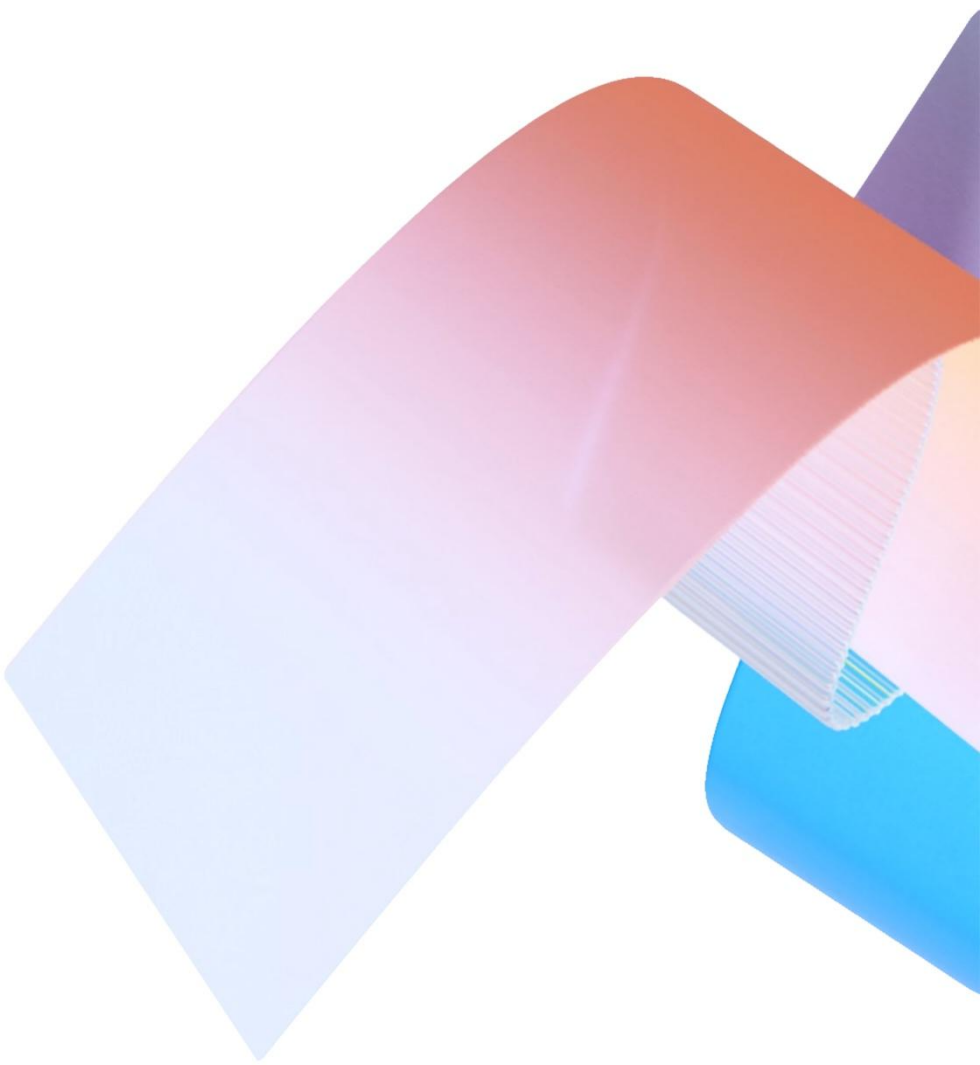
Microsoft

Administering and Governing Agents



1 Introduction

1.1 Purpose

This document aims to provide a comprehensive overview of the strategies and tools necessary to secure and govern agents within Microsoft 365 environments. By detailing the functionalities and capabilities of various agent types, it seeks to equip IT professionals with the knowledge required to effectively manage and safeguard their organizational data.

This whitepaper delves into the specifics of Microsoft's tools and methodologies, offering insights on managing data security and agent integrity.

Furthermore, this document addresses common challenges faced by IT practitioners and decision-makers, such as ensuring appropriate data access levels for agents, preventing data exfiltration, and complying with relevant regulations. By providing practical solutions and recommended practices, this whitepaper serves as a valuable resource for IT departments in both Small to Medium Businesses (SMBs) and Large Enterprises, guiding them towards a more secure and well-governed agent management strategy.

This version has been updated to include new information on zone governance strategies, sharing controls, and agent owner reassignment.

1.2 Scope

This whitepaper focuses on governing agents within the Microsoft 365 (M365) ecosystem.

- **SharePoint:** Use SharePoint to build agents based on content stored in SharePoint sites and libraries. Learn more at Get started with SharePoint agents.
- **Agent Builder in Microsoft 365 Copilot:** Use Agent Builder directly within M365 Copilot to create conversational templates tailored to specific tasks or business needs. Learn more at Use Copilot Studio to build agents.
- **Copilot Studio:** Use triggers, advanced logic, and connections to other Microsoft services or third-party platforms to create agents with Copilot Studio. Learn more at Overview - Microsoft Copilot Studio.

- **Pro developer tools:** Use tools and services like Team Toolkit and Microsoft Foundry to build fully customized agents with the model and orchestration engine of your choice.

1.3 Target audience

The target audience for this whitepaper is IT Practitioners and IT Decision-Makers in Small to Medium Businesses (SMB) and Large Enterprises who are responsible for their organization's agent management and governance strategy.

2 Overview

To establish a strong foundation for agent governance within the Microsoft 365 ecosystem, it is essential to explore core governance principles. Structured frameworks such as zones and governance pillars should also be leveraged, enabling IT practitioners and decision-makers to align their strategies with organizational goals while navigating complex challenges.

The following sections delve into governance principles and structured frameworks in more detail.

2.1 Governance fundamentals

To ensure success, it is recommended that fundamental governance principles are established. Conceptually, these principles can be summarized into three pillars that support the governance system.

Policy: There should be a comprehensive strategy outlining the system's objectives, accompanied by policies to provide high-level structure. For example, the overarching strategy may involve pilot agents to evaluate business benefits. To achieve this, policies may be required for data loss prevention, citizen development, and Application Lifecycle Management (ALM).

Process: The methodology for implementing this strategy is defined by processes. For instance, there should be a process for regular review and reporting of costs, as

well as a process for ALM. Clearly defined issue escalation paths should also be established.

People: Governance is inherently complex, and requires management by individuals who are authorized, capable, and accountable.

No matter how many governance tools are available, true governance only happens when organizations define a clear strategy and actively implement supporting processes. Tools can help automate and enforce, but without strategic intent and disciplined execution, governance remains aspirational rather than operational.

While no set of tools can replace the need for a clear governance strategy and disciplined processes, the right tools can make it much easier to operationalize and sustain those practices at scale.

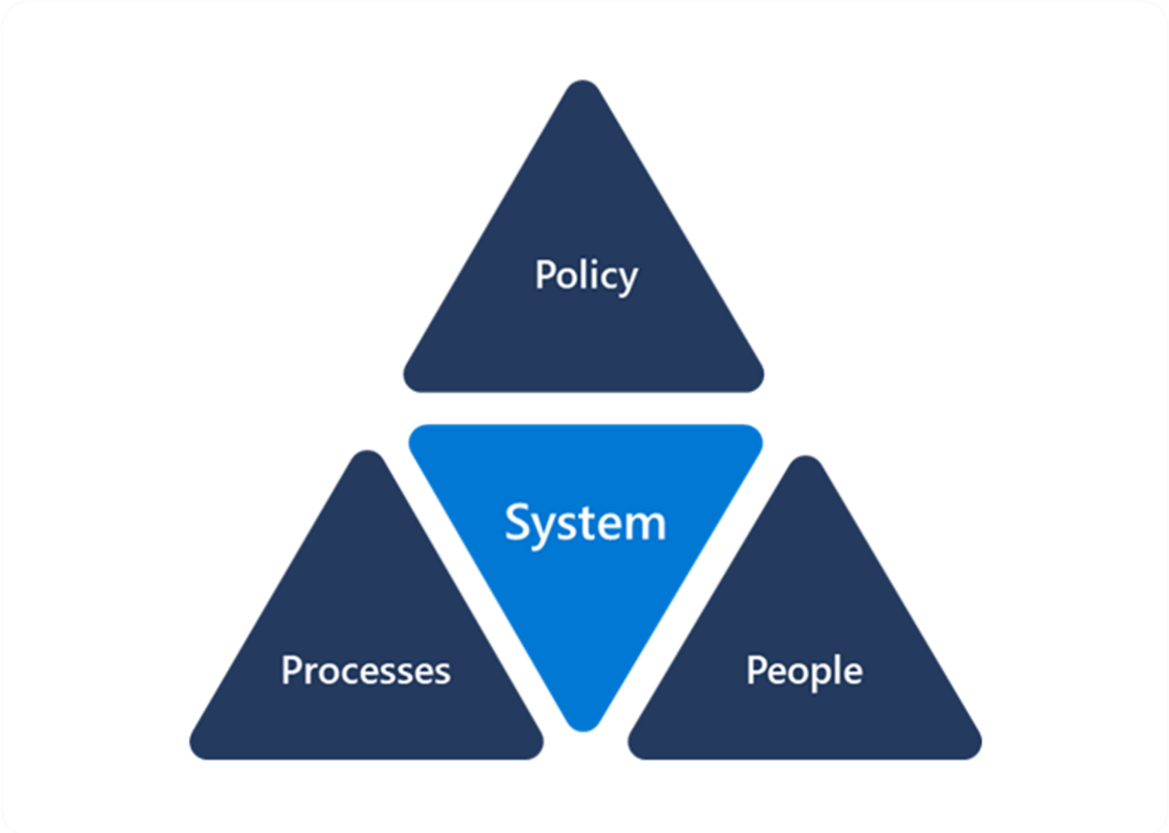


Figure 1 Governance pyramid

At a conceptual level, the pillars of governance can be presented as a pyramid. This pyramid emphasizes **the importance of non-system factors in achieving a robust governance system**.

Within the Microsoft governance ecosystem these governance pillars are enforced with Environment Groups, Group Rules and Environment Routing.

2.2 Zones

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, 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.

At a practical level, Organizations implement zones through Environment Groups with the Copilot Studio admin controls featured in the Power Platform admin center (PPAC). Environment rules prevent drift and apply consistently across grouped environments.

2.2.1 Zone 1: personal productivity

Zone 1 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.

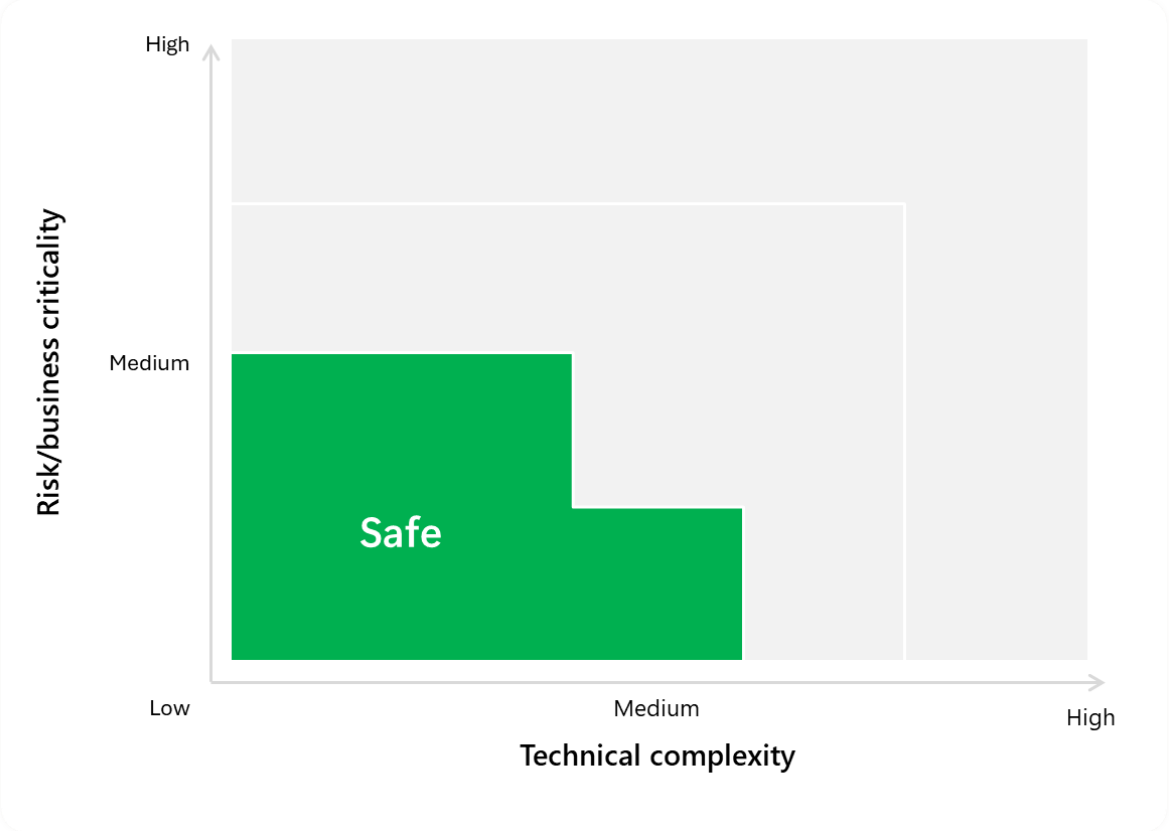


Figure 2 Personal productivity zone

2.2.2 Zone 2: team collaboration

Zone 2 builds upon zone 1 and represents a step forward in governance maturity, empowering organizations to maintain operational excellence while adapting to the complexities of modern IT environments.

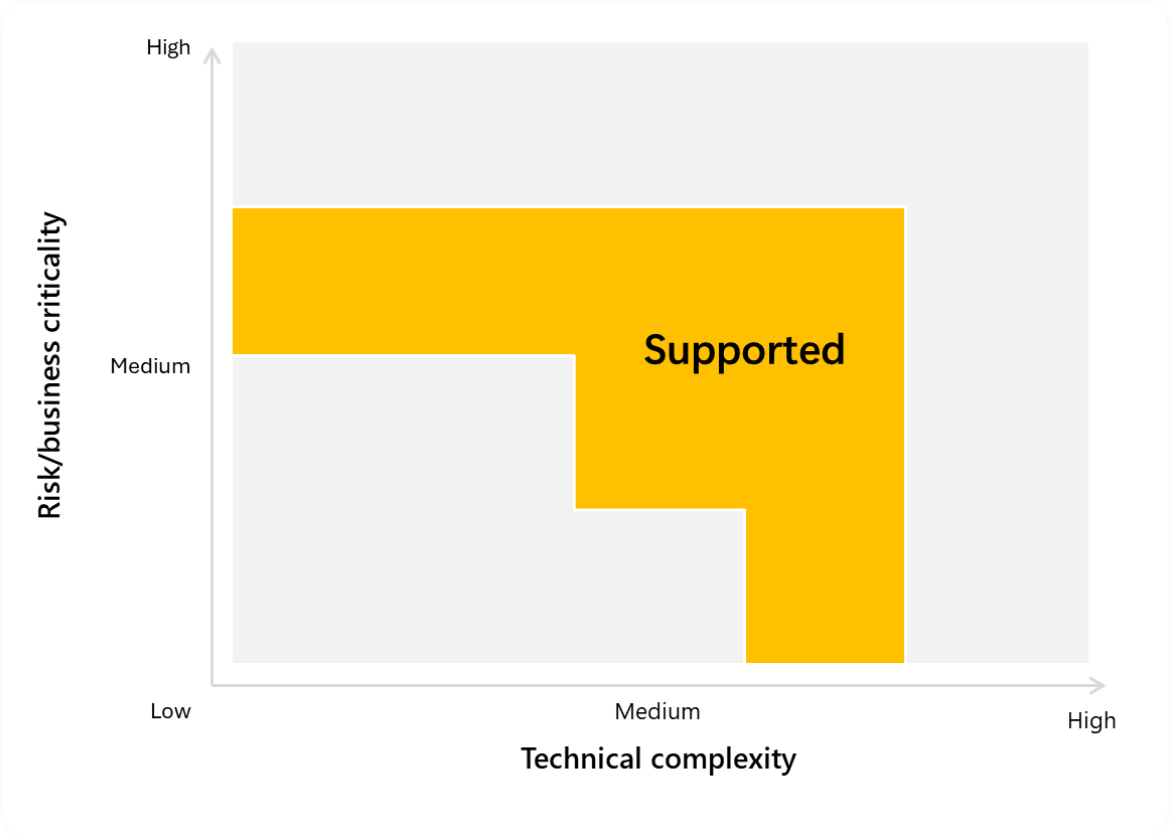


Figure 3 Team collaboration zone

2.2.3 Zone 3: enterprise managed

Zone 3 builds upon the zone 2 (and hence the zone 1) 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 and continuous monitoring, while supporting complex agent scenarios with scalable management controls and actionable reporting for strategic decision-making.

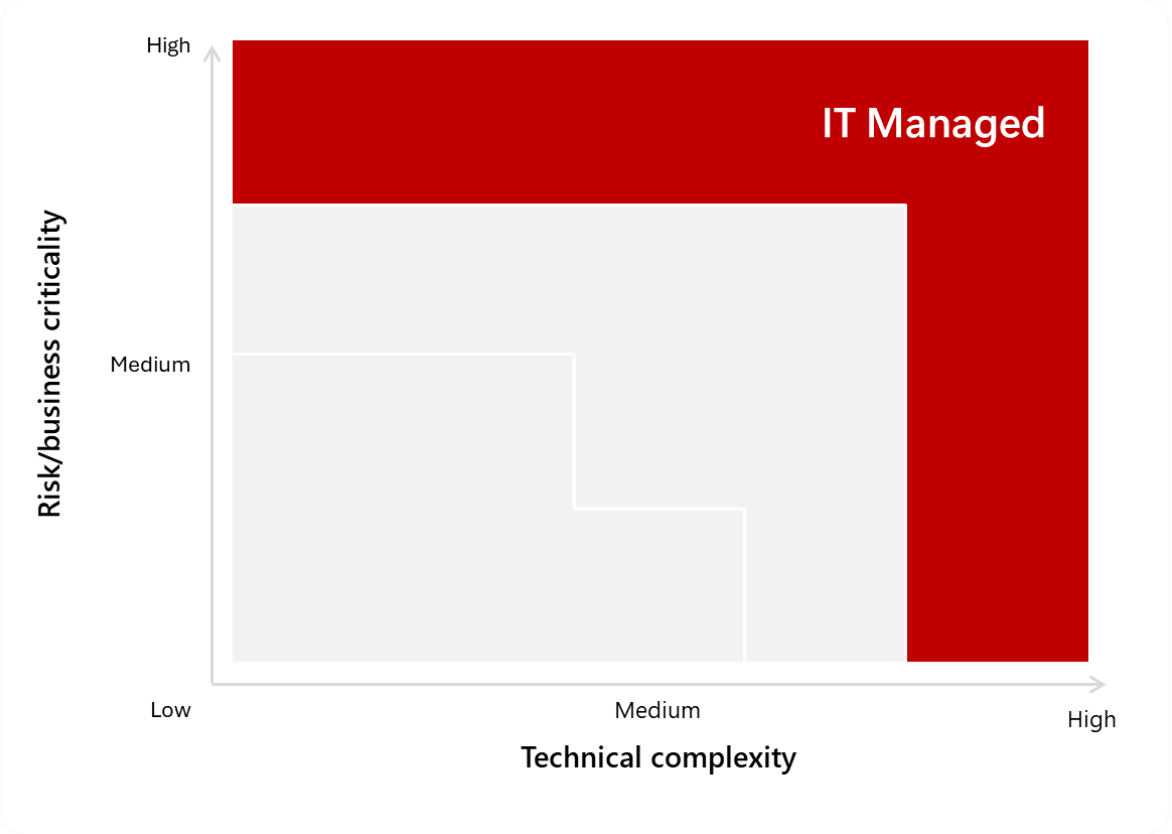


Figure 4 Enterprise managed zone

2.3 Governance pillars

Effective governance within the Microsoft 365 ecosystem requires a structured and tiered approach to security, management, and operational strategies. By segmenting governance into distinct pillars, organizations can address their evolving needs while maintaining scalability and adaptability. This section outlines the core principles of governance maturity and the framework needed to ensure operational excellence, security enhancement, and strategic alignment within complex IT environments.

2.3.1 Security controls

Security controls are measures, policies, and procedures designed to safeguard an organization's assets, data, and systems from unauthorized access, misuse, or exploitation. These controls typically encompass technical, administrative, and physical elements that work in concert to mitigate risks and ensure compliance with regulatory and organizational standards. From encryption protocols and access management to monitoring tools and incident response strategies, security controls form the backbone of a robust governance framework. Their implementation ensures that data integrity, confidentiality, and availability are maintained across all zones and operational layers.

2.3.2 Management controls

Management controls refer to the set of policies, procedures, and operational guidelines that organizations implement to oversee, direct, and ensure the effective functioning of their agents and services within the Microsoft 365 ecosystem. These controls provide a structured framework for managing the lifecycle of agents, including their creation, deployment, maintenance, and decommissioning.

By establishing clear roles and responsibilities, IT administrators can maintain consistency in configurations, enforce compliance with organizational standards, and streamline workflows. Management controls often involve processes such as role-based access management, policy enforcement, resource allocation, and periodic audits to verify adherence to security and operational protocols. In addition, these controls help ensure seamless integration within the broader IT architecture, facilitating collaboration and scalability across different zones of governance.

2.3.3 Agent reporting

Agent reporting is the process of systematically collecting, analyzing, and presenting data related to the activities, performance, and compliance of agents within an organization. This reporting ensures transparency and accountability by providing insights into how agents interact with the system and contribute to operational objectives. It includes metrics such as agent deployment status, access control adherence, system interactions, and error logs. By leveraging reporting tools, IT administrators can identify trends, track anomalies, and address issues proactively, enabling better decision-making and resource optimization. Agent reporting serves as a critical element within the governance framework, bridging the gap between

security controls and management controls while offering actionable intelligence for maintaining operational excellence.

The new reporting architecture delivers this visibility primarily through four key experiences with the Copilot Studio admin controls: Inventory, Monitoring, Security, and Copilot. In addition, Microsoft Purview provides the cross-tenant data security and compliance reporting layer - for example, Data Loss Prevention (DLP), Data Security Posture Management for AI, and Insider Risk - which surfaces how agents interact with sensitive data across Microsoft 365. Later sections in this paper describe how PPAC and Microsoft Purview work together to deliver an end-to-end reporting story.

2.3.3.1 Inventory – Tenant-wide visibility

The Inventory experience provides a comprehensive, cross-environment snapshot of every app, flow, and Copilot agent within the tenant. As of October 2025, it is in public preview. It enables admins to see what exists where, who owns it, and how it is being used, forming the baseline catalog for governance decisions.

2.3.3.2 Monitoring – Health and performance insights

The Monitoring experience is available with the Copilot Studio admin controls. It offers operational analytics for Copilot Studio agents. It tracks session success rates, highlights degradation trends, and can raise alerts when an agent's performance dips or fails. At the time of writing, this capability is in public preview.

2.3.3.3 Security – Centralized posture management

The Security experience is part of the new reporting architecture within the Copilot Studio admin controls. It centralizes the tenant's security posture across Copilot Studio, including data access, connector configuration, and threat-related recommendations. It surfaces insights and remediation guidance that strengthen Copilot Studio security posture and helps administrators quickly identify misconfigurations or risky connectors.

2.3.3.4 Copilot – Governance and ROI command center

The Copilot experience consolidates governance, analytics, and business value metrics for Copilot usage. It acts as a command center for agents, providing:

- Usage and ROI dashboards to justify spend and adoption trends.

- Centralized governance controls for sharing, Purview DLP, and advanced connector policies.
- Innovation planning insights and recommendations tailored to each environment's maturity.

For many organizations, the Copilot experience becomes the executive-facing reporting layer that links technical compliance with business outcomes.

2.3.4 Maker routing and environment auto-provisioning

To ensure makers always build agents in the correct governance zone, Microsoft recommends using **Environment Groups with Routing**. Routing automatically places each maker into the appropriate Copilot Studio environment based on organizational rules.

When a user opens Copilot Studio or Power Apps for the first time, routing evaluates their identity, role, or security group membership and provisions them into:

- **Zone 1 – Personal Developer Environment** (an auto-provisioned, isolated development environment — **not** the Default Environment).
- **Zone 2 – Team Collaboration Environment** (a managed departmental environment for shared agents).
- **Zone 3 – Enterprise Managed Environment** (for high-risk or organization-wide agents).

Routing ensures that makers cannot accidentally create agents in an environment with incorrect connector policies, sharing defaults, or data access rules. It also discourages "shadow AI" by guaranteeing that every new agent is created inside a zone that already enforces:

- Connector management policies (ACP)
- Sharing and publishing restrictions
- Publishing controls
- Security baselines
- Environment ownership and support channels

Routing is configured by administrators Copilot Studio admin controls and forms a foundational component of zoned governance—ensuring the development lifecycle begins in the right place before any additional controls are applied.

2.3.5 New and preview controls

As growth of agents across the workforce has been rapid, the need to effectively govern them securely is increasingly a salient challenge. To this end Microsoft has recently released several tools to assist in the management of agents.

Several noteworthy examples are:

- Rules (connector policies, Agent AI Models, Sharing, Channels, authentication and onboarding): Learn: Rules for environment groups
- Default environment routing: Learn: Environment routing
- Management controls for developer environments: Learn: Environment management capabilities
- Power Platform connectors: Learn: Power Platform for Admins V2
- Microsoft responsible AI: Learn: Microsoft Responsible AI Tools and Guidelines
- Power Platform Inventory: Learn: Power Platform inventory (preview)
- Copilot Adoption in Power Platform: Learn: Track, manage, and scale Copilot adoption
- Enhancements to the pipeline configuration tool: Learn: Set up pipelines
- Power Platform adoption guidance documentation: Learn: Microsoft Power Platform guidance documentation
- Microsoft Power Platform and Copilot Studio Architecture Center: Learn: Microsoft Power Platform and Copilot Studio Architecture Center
- Copilot studio kit: Learn: Copilot Studio Kit overview
- Health monitoring and KPI Tools: Learn: Monitor the health of your resources
- Additional Runtime Protection: Learn: Enabling external threat detection and protection for Copilot Studio custom agents (preview), Learn: Protect your environment in real-time during agent runtime

These tools provide a wide range of enhancements to the existing suite of products, providing even greater oversight and monitoring capability.

Given the pace of change it is advised that governance practitioners stay up-to-date with the latest information: Key concepts - Copilot Studio security and governance

2.4 Discovery and entry point for agents

The Microsoft 365 Copilot Agent Store is a primary discovery and access experience for agents within Microsoft 365 ecosystem. It provides users with a designed experience intended to provide a consistent way to find, evaluate, and install agents aligned to their role, task, or workflow. From a governance perspective, it creates a

controlled entry point where only agents that meet organizational, security, and compliance requirements are made available for use.

Agents in the store fall into three categories.

- Microsoft-built agents are first-party agents developed and maintained by Microsoft.
- Partner-built agents are third-party agents published to the store by software development companies.
- Customer-built agents are those created by your own organization and published to the store, including agents built using Microsoft Copilot Studio, Microsoft Foundry, or an external platform.

Additionally, agents built with Agent Builder and shared by colleagues appear in the recipient's **Shared with me** collection, while Agent 365-enabled agents appear in the **Agents for your team** collection within the store.

By serving as the governed entry point for agent discovery, the Agent Store can help organizations scale agent adoption safely and intentionally, balancing innovation and accessibility with the security, oversight, and control required in your environment.

3 Zone 1: personal productivity

Zone 1 represents the foundational level of governance within an organization's agent management and security framework. It is characterized by high-level configuration settings designed to establish baseline security and management protocols.

As noted previously, zones are a conceptual framework and are practically implemented as environment groups.

These settings ensure the implementation of essential security controls and management controls to safeguard organizational assets and streamline agent operations.

The key features are:

- **Security controls:** This zone emphasizes basic security measures such as encryption protocols, access management, and incident response strategies to maintain data integrity, confidentiality, and availability.
- **Management controls:** It incorporates policies and procedures for managing the lifecycle of agents, including their creation, deployment, maintenance, and decommissioning. Processes like role-based access management and policy enforcement are core to this zone.
- **Agent reporting:** Zone 1 involves collecting and analyzing data on agent performance and compliance, ensuring transparency and accountability within the governance framework. Metrics like deployment status and system interactions are crucial components.
- **Environment groups:** These are the tools used to apply consistent governance across all environments including those in zone 1. This zone includes all developer environments enabling governance to be applied across the full agent lifecycle.

3.1 Agent characteristics

Agents in the personal productivity zone are designed specifically to assist with personal tasks. These agents are characterized as low-risk entities that primarily interact with data from Microsoft Graph. The environment within this zone is tailored for developers and includes limited connectors. Notably, these agents are not shared and are confined to individual use, ensuring a secure and isolated operational framework. The focus of the personal productivity zone lies in empowering individuals to leverage Copilot Studio agents for streamlined and efficient management of personal productivity tasks while maintaining a controlled and risk-minimized environment.

The environments in which these agents operate are admin provisioned. This means that administrators are responsible for setting up and managing these environments, ensuring that they align with organizational needs and policies.

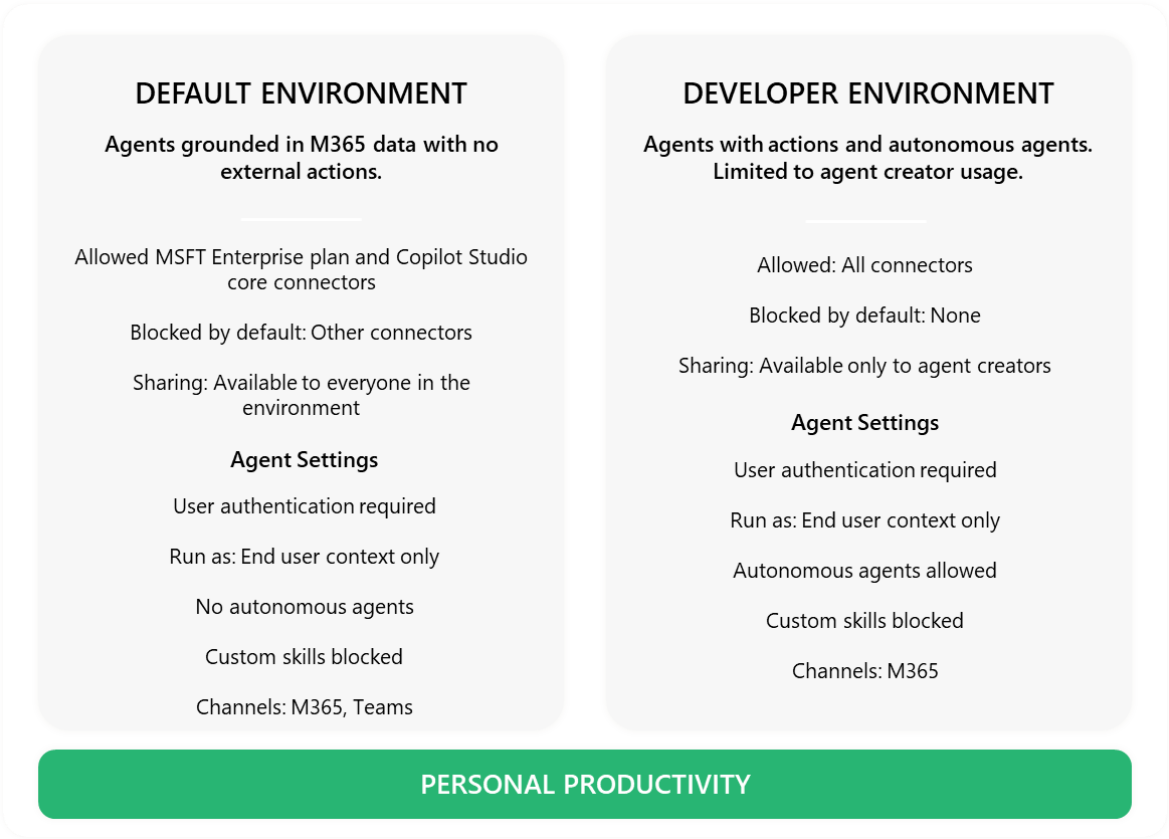


Figure 5 Zone 1 - default configuration

3.2 Security controls

3.2.1 Microsoft 365 admin center

The Microsoft 365 admin center (MAC) is a portal that offers management and governance for Microsoft 365 services. It allows IT administrators to monitor, configure, and manage their environment, ensuring performance, security, and compliance.

Administrators can manage user accounts, assign licenses, and control access to applications through the MAC. This simplifies user management and maintains organizational consistency. The MAC also provides analytics and reporting tools for insights into user activity, system performance, and security threats.

It integrates with other Microsoft services like Azure for enhanced functionality, such as identity and access management. By using the MAC, organizations can streamline operations, reduce administrative tasks, and ensure security and compliance.

The MAC governs both End-User created agents and IT Catalog agents.

Learn more about managing Copilot at Manage Microsoft 365 Copilot scenarios in the Microsoft 365 admin center.

3.2.2 SharePoint Online content permissions

SharePoint agents are managed through the existing robust SharePoint Online permission system that ensures only authorized users can interact with sensitive data and perform specific tasks. SharePoint's permissions are structured into distinct levels, such as Full Control, Edit, Contribute, and Read, each providing a varying degree of access and control over the content and settings.

This structure ensures that users have the necessary access to perform their tasks while maintaining the site's security and integrity.

Therefore, SharePoint agents can only access content or update data that the current user already has access to, because SharePoint agents are themselves stored as files. These permissions can also govern user access to SharePoint agents.

This hierarchical structure of permissions helps maintain a secure environment where SharePoint agents can operate efficiently, minimizing the risk of unauthorized access or changes. It also ensures that users have the appropriate level of access needed to perform their tasks without compromising the integrity or security of the SharePoint site.

Learn more about agents and SharePoint Online content permissions at Manage access to SharePoint agents - SharePoint in Microsoft 365.

3.3 Management controls

3.3.1 SharePoint Advanced Management

SharePoint Advanced Management (SAM) allows users to control Copilot behavior at the content source level. The Restricted Content Discovery policy hides a site's content from global searches and Copilot indexing, protecting sensitive data. SAM includes tools like site sharing restrictions and block download policies to prevent oversharing of confidential information and can be used to detect and remediate oversharing of content.

With SAM, administrators can tailor the accessibility and visibility of content within SharePoint, ensuring that only authorized users have access to specific data. The

feature set helps maintain compliance with data protection regulations by limiting exposure of sensitive information. For example, the site sharing restrictions enable administrators to specify who can share content and with whom, thereby reducing the risk of data breaches. Additionally, the block download policies prevent users from downloading files from certain sites, which is especially useful for protecting proprietary or confidential documents. By leveraging these advanced management tools, organizations can enhance their data security posture while still enabling collaboration and productivity.

Learn more about SharePoint Advanced Management at Microsoft SharePoint Premium - SharePoint Advanced Management overview.

3.3.2 Agent sharing

Zone 1 is primarily intended for personal productivity and Personal Developer environments. Agent sharing in zone 1 is generally constrained to maintain a secure, personal workspace for each maker but can be allowed by administrators with the following confines:

- Sharing disabled entirely,
- Sharing enabled for specific security groups, or
- Sharing enabled with viewer-only access (editor access does not apply to personal productivity scenarios).

By default, Agent Builder in Microsoft 365 Copilot environments are provisioned as personal developer environments, meaning that sharing may be disabled entirely or limited to a single additional person for peer review or demonstration purposes.

It is important to distinguish between sharing and publishing. Sharing determines who is allowed to access the agent. Publishing determines *how* the agent is exposed through channels such as Teams, websites, or other integrations. External channels (e.g., WhatsApp, public webchat) relate to publishing—not sharing—and are normally blocked in personal environments through organization-level publishing policies.

If an agent needs to be shared more broadly—for example, beyond a small security group or with a functional team—it should be promoted to zone 2 (team collaboration) and deployed in a managed environment. Zone 2 environments enforce connector management policies via environment group rules, ensuring that only approved connectors and data sources are available when agents are shared

with larger audiences. Pipelines or solution-based ALM processes may also be used to promote agents between environments for consistency and security.

These controls minimize the risk of oversharing or data leakage while allowing makers to innovate safely within their own isolated sandbox. They also help ensure that only authenticated users can interact with agents and that any growth in usage follows the organization's structured governance path from zone 1 → zone 2 → zone 3 zones.

Learn more about agent sharing at Share and manage agents.

Learn more about disabling or limiting Copilot Studio agents at Control how agents are shared.

Learn more about limit sharing at how limits can be imposed on sharing.

3.4 Agent reporting

3.4.1 Agent inventory

All agents are treated as apps in the system, allowing administrators to have a centralized inventory. The Integrated Apps page of the Microsoft 365 admin center (MAC) lists all apps as well as all agents available in the tenant, enabling actions like blocking agents or assigning them to specific users. Administrators maintain full app management capabilities for agents, ensuring a catalog of allowed agents and intervening when necessary.

The available reports provide an overview of the agents that the organization has developed using various tools such as Copilot Studio, Agent Builder, and Teams Toolkit. This comprehensive view allows administrators to see all the agents created under the organization's umbrella.

Learn more about inventory management at Get started with Integrated apps - Microsoft 365 admin.

Learn more about agent inventory at Manage agents for Microsoft 365 Copilot in Integrated Apps - Microsoft 365 admin.

4 Zone 2: team collaboration

Zone 2, the team collaboration zone, builds upon the foundational elements of zone 1 and introduces advanced governance and operational protocols to enhance an organization's security and management framework. It is designed for organizations that require stricter enforcement of policies, more granular configuration settings, and deeper integration of reporting mechanisms within the Microsoft 365 ecosystem.

The key features are:

- **Security controls:** Zone 2 environments enforce stricter data-access and connector controls than zone 1. Environment Group rules apply connector management policies, limiting which connectors, APIs, and data sources can be used when agents are created or shared. Sensitive or departmental data sources may be allowed, while external or high-risk connectors remain restricted. Access to the environment is scoped to specific Microsoft Entra security groups, ensuring that only approved makers and team members can create, modify, or run agents.
- **Management controls:** This zone focuses on detailed lifecycle management of agents, including automated workflows for deployment and decommissioning, stricter role-based access controls, and continuous policy updates to adapt to organizational needs.
- **Agent reporting:** Reporting capabilities in zone 2 are more robust, offering granular insights into agent behaviors, compliance metrics, and system anomalies. These enhanced reports enable IT administrators to make informed decisions and optimize resource utilization.

4.1 Agent characteristics

Agents in the team collaboration zone are specifically designed to support team-level solutions that are integral to organizational processes. Their role is to streamline collaborative efforts, ensuring operational efficiency across departments—and provide agentic efficiency to new or existing organizational processes.

These agents include tools like HR solutions that automate human resource-related functions, making tasks such as onboarding smoother and more efficient. Another example is team reporting agents, which assist in generating and analyzing reports to enhance decision-making for teams.

As with the zone 1, the environments in which these agents operate are admin provisioned. This means that administrators are still responsible for setting up and managing these environments, ensuring that they align with organizational needs and policies.

Security plays a critical role in the team collaboration zone. These agents are equipped with robust advanced connector policies to protect sensitive information. Additionally, scoped roles ensure that access to data and system features is carefully controlled, providing a tailored approach to user permissions based on their specific responsibilities within the organization.

The team collaboration zone integrates sophisticated governance mechanisms to maintain compliance and operational integrity. These agents are optimized for scalability and adaptability, empowering teams to dynamically respond to evolving challenges while safeguarding organizational resources.

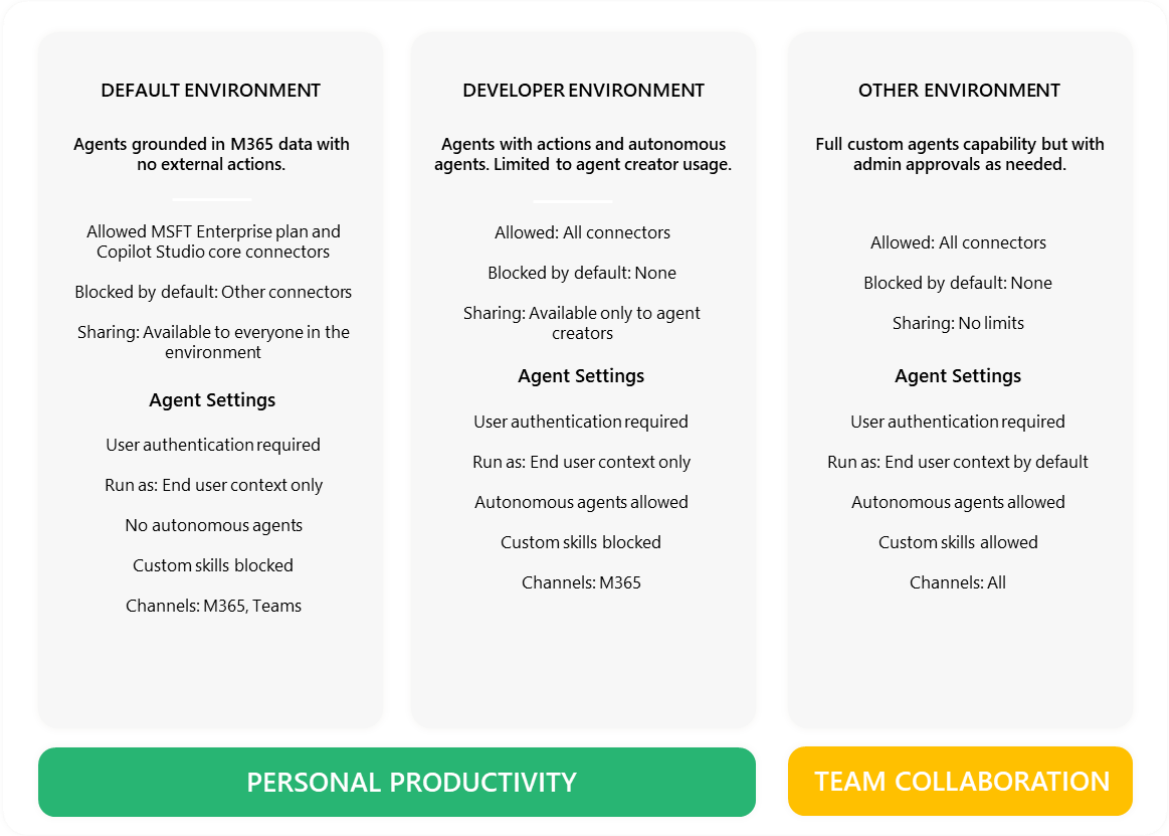


Figure 6 Zone 2 default configuration

4.2 Security controls

4.2.1 Copilot Studio admin controls

The Copilot Studio admin controls within the Power Platform admin center (PPAC) serves as one of the central portals for governing Copilot Studio, empowering administrators to oversee and secure AI-powered agents created by agent creators. Through robust controls, administrators can define and enforce policies that dictate agent behavior, ensuring compliance while maintaining flexibility for innovation. Additionally, granular governance features allow administrators to monitor and restrict unauthorized actions, ensuring agents operate securely within pre-determined organizational boundaries.

Microsoft Copilot Studio agents adhere to organizational content governance through Microsoft Purview: sensitivity labels that prevent agents from processing files; data protection policies; agent rules; data masking; geographic boundaries; and agent interface boundaries. Data masking hides sensitive data elements, allowing authorized users to perform their tasks without exposing the actual data.

Cost management is another key feature available for agent governance within the Copilot Studio admin controls, in that administrators can enable a billing plan to monitor consumption of agents and agent messaging.

By implementing these measures, the Copilot Studio admin controls ensures a robust framework for managing and protecting organizational content, maintaining compliance with regulatory standards, and safeguarding against potential risks. As such, it provides governance for maker-created agents.

Learn more about how to Create and manage masking rules (preview).

Learn more about data storage and governance at Data storage and governance.

Learn more about how to Set up a pay-as-you-go plan.

Learn more about using the Copilot Studio admin controls with the Power Platform admin center at Overview of the Power Platform admin center.

4.2.2 Copilot Studio data policies

Administrators can govern how agents interact with data and services using advanced connector policies in the Copilot Studio admin controls. These policies

allow administrators to control which connectors are allowed or blocked, ensuring agents comply with organizational security and compliance requirements.

Advanced connector policies can be used to:

- **Classify and control** connectors used by Copilot Studio resources.
- **Restrict** high-risk or external connectors and define which internal data sources are in-scope.
- **Apply** different connector rules per environment group (for example, more permissive rules in a development zone and stricter rules in production).
- **Limit** or **block** specific actions within connectors where supported, not just the connector as a whole.

These policies are configured with the Copilot Studio admin controls and apply at tenant and environment levels. They complement—but do not replace—Microsoft Purview DLP, which governs how sensitive data can be used and shared across the broader Microsoft 365 estate.

Learn more about advanced connector policies at Advanced connector policies.

Learn more about managing data policies at Manage data policies.

4.3 Management controls

4.3.1 Copilot Studio environment controls

Copilot Studio environments provide a secure and organized way to manage apps, flows, and data by separating development, testing, and production work. They support governance through role-based access, advanced connector policies and data policies, and solution-based application lifecycle management. This structure helps teams scale effectively, maintain control, and deploy changes safely across different parts of the organization.

Environment controls are inherited by Copilot Studio admin controls and are required to maintain secure and efficient operations. These controls enable administrators to define and enforce policies that govern the use of resources, data connections, and user access across different environments. By establishing clear guidelines and restrictions, Copilot Studio environment controls help prevent unauthorized activities and ensure compliance with organizational standards. These controls extend to each component in the agents and agent services.

One of the primary features of these controls is their ability to monitor and manage data flows between various applications and services, thereby safeguarding sensitive information and preventing data breaches. Additionally, they provide tools for auditing the activities of administrators, agent creators, and users, as well as tracking changes within the environment. These capabilities are crucial for maintaining transparency and accountability.

The key controls that should be implemented include:

- **Environment types:** Developer, Sandbox, Trial, Production, etc.
- **Security controls**: Roles, security groups, advanced connector policies, and sharing settings.
- **Data management controls:** Dataverse access settings, storage capacity, backup/restore options, and reset functionalities.
- **Deployment controls:** Application Lifecycle Management (ALM), pipelines, and version control integration.

Learn more about environment controls at Managed Environments overview.

4.3.2 Authoring agents

Admins can effectively provide the toolset for agent creators to create and author agents by creating environments and environment scopes, setting environment roles, from within Copilot Studio admin controls. This can include a welcome message that informs agent creators and Copilot Studio agent creators of the policies and practices of the environment that have been shared with them.

Organizations should ensure that each Copilot Studio agent creator is assigned a personal development environment to support independent and efficient work with necessary resources and tools. While the assignment of environments establishes ownership and access, the provisioning of these environments is best handled through environment routing[1].

Admins can also export and import agents using solutions – enabling them to be moved around multiple environments.

Learn more about enabling a welcome message at Enable maker welcome content.

Learn more about Export and import agents using solutions - Microsoft Copilot Studio.

[1] Environment management and routing is covered in *Section 4.3.3 Environment management*.

4.3.3 Environment management

Managed environments allow organizations to implement guidelines and standards at scale when generating new environments. This includes adhering to security protocols, software licensing agreements, and organizational policies to ensure compliance and consistency across all environments.

Environment management in the Copilot Studio admin controls governs both agent creators and the agents they create – ensuring innovation stays within guardrails. Administrators use security groups and rules to control who can modify artifacts within environments (and the environments themselves), preventing unauthorized changes to Copilot Studio agents, chatbots, or workflows, while agent creators (including citizen developers) can freely build within compliance established by policy. This balances flexibility with security: agents run as intended, sensitive data stays protected, and no single agent creator can override organizational standards. Ensuring agent creators can build agents in personal environments is also a way to ensure other agent creators don't access their agents during development and testing.

In addition, Pipelines can be configured based on an already approved environment strategy. It is recommended that administrators familiarize themselves with Pipelines.

Once the Environments have been created, administrators can use the Copilot Studio admin controls Deployment page to create and establish Pipelines.

Once a Pipeline(s) has been established, agent creators can use Pipelines to deploy and certify agent solutions from development to production in a repeatable, governed, and secure manner. Pipelines automate key steps such as solution packaging, environment configuration, and deployment, reducing manual effort and the potential for errors. By integrating approval workflows, organizations can also introduce a "human in the loop" to review and validate agents before they go live, ensuring quality and alignment with business goals. This structured approach enhances visibility, enforces consistency, and accelerates the overall delivery lifecycle.

Learn more about the environment types and their use cases at environments overview.

Learn more about environment routing at environment routing.

Learn more about Pipelines at Overview of pipelines.

4.3.4 Agent publishing controls and permission updates

To prevent unauthorized changes to agents after updates have been completed, the publish feature can be disabled through connector management. By doing so, admins ensure that no further modifications can be made without proper authorization, thereby safeguarding the agent's content from potential tampering.

Admins should regularly update permissions to identify and mitigate vulnerabilities, thus maintaining system security. This involves periodically reviewing who has access and removing unnecessary permissions.

Learn more about agent publishing controls and permission updates at Key concepts - Publish and deploy your agent.

4.3.5 Agent sharing

Sharing in zone 2 enables small teams or departments to use and co-author Agent Builder in Microsoft 365 Copilot agents within a managed environment. Sharing settings are controlled by administrators and typically allow access for:

- **Specific Microsoft Entra security groups** (e.g., HR Analysts, Finance Team)
- **Named users**, or
- **Viewer-only access** where teams can run the agent but not modify it.

Editor access can be granted in zone 2 when the environment is intentionally configured to support collaborative development. Environment Group rules enforce the zone's connector management policies, ensuring that only approved data sources, actions, and connectors can be used when agents are shared.

It is important to differentiate **sharing** from **publishing**. Sharing grants users' permission to access the agent within the environment. Publishing determines how the agent is surfaced—such as making it available in a Teams channel, embedding it in a departmental site, or exposing it as a plugin to other M365 experiences. Publishing options available in zone 2 are controlled by administrators and typically limited to **internal channels only**, never external.

If the agent attracts a wider audience, requires elevated data access, or becomes business-critical, it should be **promoted to zone 3** and deployed within an enterprise-managed environment with additional controls and operational oversight.

These sharing controls protect team-level data, prevent oversharing, and ensure that all use remains within the department's governance boundary while still enabling collaborative innovation.

4.4 Agent reporting

4.4.1 Microsoft admin center Integrated Apps

Refer to the zone 1 Section **Error! Reference source not found. Error! Reference source not found.**.

4.4.2 Microsoft Purview

Microsoft Purview delivers unified data security, governance, and compliance solutions designed to protect and govern the organization's data estate in the era of AI. As organizations increasingly rely on AI to streamline operations and enhance decision-making, ensuring robust data security becomes paramount.

Learn more about Microsoft Purview at Microsoft Purview.

4.4.2.1 Data Security Posture Management for AI

Data Security Posture Management (DPSM) for AI enables customers to gain visibility into agent usage and assess data risks, including:

- **Discover sensitive data shared in AI agent interactions:** Customers can track the sensitive information used for grounding and get recommended actions on quick one-click policies to improve their data security posture.
- **Detect risky AI usage:** Customers can explore risk analytics and investigate risky user activities, such as a departing employee having an unusual number of prompts with sensitive data.
- **Regulatory compliance:** Customers can also explore unethical AI interactions such as targeted harassment or unauthorized disclosures, ensuring they maintain regulatory compliance.

A data security administrator can review agent prompts, responses, and sensitive grounding data within the DSPM for AI activity explorer. This role involves monitoring and analyzing how AI systems interact with users and handle data to ensure compliance with security policies and regulations. By doing so, the administrator can identify potential vulnerabilities or misuse of data, investigate, and

implement measures to mitigate risks. Additionally, they should work closely with other IT and security teams to develop strategies for protecting sensitive information and maintaining the integrity of AI operations.

A data security administrator can also identify data at risk of being excessively shared by agents when using SharePoint grounding data. This involves understanding access patterns, usage behavior, and sharing activities to detect any potential vulnerabilities or breaches. They may also enforce policies to mitigate risks and maintain the integrity and confidentiality of the data.

Learn more about data security posture management for AI at Considerations for deploying Microsoft Purview Data Security Posture Management for AI to manage data security and compliance protections for AI interactions.

4.4.3 Copilot Studio admin controls

The Copilot Studio admin controls offers comprehensive tools for tracking and managing agents created within Microsoft Copilot Studio. By serving as a central governance portal, it empowers administrators to monitor agent activity and ensure compliance with organizational policies.

Agent behavior policies allow administrators to define rules that govern how agents behave, ensuring alignment with organizational and regulatory standards. Advanced connector policies and data policies enable administrators to classify and safeguard access to data by controlling which connectors—and in some cases which connector actions – can be used in each environment or environment group. This reduces the risk of data exfiltration while preserving productivity.

Granular governance features in PPAC help administrators monitor unauthorized actions and impose restrictions to ensure agents operate securely within organizational boundaries. Content governance itself is supported by Microsoft Purview (for example, sensitivity labels and DLP policies), which PPAC honors when agents work with Microsoft 365 content such as SharePoint grounding data.

Additionally, cost management is facilitated by the billing plan feature, enabling administrators to track agent consumption and messaging, thereby providing insights into operational costs and usage patterns.

By leveraging PPAC's tracking and governance capabilities alongside Microsoft Purview, organizations can maintain compliance with regulatory standards, protect sensitive data, guard against unauthorized actions or breaches, optimize cost

monitoring and resource allocation, and ensure agents are adaptable and secure while fostering innovation.

5 Zone 3: enterprise managed

Zone 3, the enterprise managed zone, represents the advanced stage within the governance framework, where organizations focus on optimizing and elevating their IT operations through enhanced security measures, sophisticated management protocols, and comprehensive reporting capabilities. At this level, organizations aim to achieve resilience, adaptability, and innovation within their processes.

Zone 3 emphasizes proactive governance solutions, incorporating predictive analytics, automated workflows, and integration with emerging technologies to address the dynamic challenges posed by modern IT environments.

By leveraging zone 3's capabilities, organizations can position themselves at the forefront of operational excellence while safeguarding their assets in an increasingly complex digital landscape.

The key features are:

- **Security Controls:** This zone often includes advanced threat detection systems, multi-factor authentication, and continuous monitoring for vulnerabilities.
- **Management Controls:** This zone is refined to support more complex agent scenarios, ensuring seamless scalability and efficient resource allocation.
- **Agent Reporting:** Reporting capabilities in zone 3 evolve into a predictive and prescriptive model, providing actionable intelligence that enables strategic decision-making and fosters continuous improvement.

5.1 Agent characteristics

Agents within the enterprise managed zone are specifically designed to handle sensitive or organization-wide use cases. These scenarios often require stringent control measures to ensure data integrity, security, and compliance with organizational standards. Common applications of these agents include customer

support copilots, finance agents, and API integrations. Their functionalities are tailored to meet the complex and high-demand requirements of these domains, providing reliable and secure solutions for sensitive operations.

Enterprise managed zone agents typically operate within Application Lifecycle Management (ALM) governed environments. This ensures structured development, deployment, and maintenance processes, enabling consistent performance and adherence to industry recommended practices. These agents undergo rigorous compliance reviews to ensure that they meet regulatory requirements and organizational policies. This process guarantees that they function securely and ethically within the stipulated guidelines.

5.2 Security controls

5.2.1 Microsoft Agent 365

To help organizations stay ahead of emerging threats, Microsoft Agent 365 (A365) includes Microsoft Defender protections like runtime threat protection, investigation, and hunting for agents that use the Agent 365 tools gateway to help organizations detect, block, and investigate malicious agent activities (Public Preview in May). This builds upon our announcements at Ignite, which include security posture management for Foundry and Copilot Studio agents (Public Preview in April) to proactively detect and mitigate specific AI vulnerabilities and threats in agents. Plus, detection, investigation and response for Foundry and Copilot Studio agents (Public Preview in May) will assist in investigating and remediating attacks that target agents if they occur.

Agent 365 extends the management, security, and governance processes organizations already use for employees to agents, so they can stay in control as agents become part of daily work. To find out more on the Security value included in Agent 365 and Microsoft 365 E7, see the Microsoft Security blog.

5.2.2 Microsoft Purview

Microsoft Purview delivers unified data security, governance, and compliance solutions designed to protect and govern the organization's data estate in the era of AI. It provides advanced capabilities for classifying, labeling, and managing data sensitivity, ensuring that organizations can proactively protect their sensitive information and mitigate any data oversharing or leakage risks. By leveraging over

315 out-of-the box and machine learning-driven classifiers, Purview can automatically detect and categorize sensitive information in agent interactions, such as personally identifiable information (PII) and financial records—to prevent unauthorized access and mitigate potential data breaches.

Additionally, Purview offers detailed risk analytics into user activity with agents, risky AI usage, and sensitive data detection, enabling organizations to make informed decisions about data security. With its extensive suite of tools, administrators can enforce data protection policies, detect unusual data access patterns, and respond to security incidents in real-time. This holistic approach ensures that all data interactions are secure and compliant with industry standards.

Data security and data compliance controls from Purview extend to SharePoint agent, Agent Builder agent, and Copilot Studio agent interactions. When agents use SharePoint as their grounding data, pre-existing sensitivity labels and data protection policies are honored.

Learn more at Microsoft Purview.

5.2.2.1 Data Loss Protection

Data Loss Prevention (DLP) helps mitigate data exfiltration risks in agents using SharePoint grounding data. Data security administrators can establish a DLP policy to restrict agents from processing files that have been marked with specific sensitivity labels if they use SharePoint grounding data.

Sensitivity labels are used to protect sensitive content. By assigning these labels to documents and files, organizations can indicate the sensitivity of their contents and ensure that only authorized personnel and systems have access to them. In the context of agents, this means that if a document is labeled as Highly Confidential (or any other label specified by the administrator), Microsoft Purview enforces the admin-authored DLP policy preventing the agent from accessing or processing Highly Confidential data and ensures that the data is not used by the agent.

This approach helps to prevent inadvertent data leaks, unauthorized access to critical information, and use of sensitive data in agent grounding data. Administrators can define rules within their DLP policy to enforce these restrictions, thus adding an extra layer of security.

Users are notified that labeled content cannot be processed due to the DLP policy in place.

Learn more about Purview DLP for Microsoft 365 Copilot.

5.2.2.2 Oversharing assessments

Purview data risk assessments for oversharing are designed to help identify and mitigate the risks associated with sharing sensitive information inappropriately. These assessments use advanced analytics to assess data sharing activities across the organization's copilots, agents, and AI applications that use SharePoint grounding data, to detect potential oversharing risks. A default data risk assessment automatically runs weekly for all SharePoint sites used by agents in the organization, or administrators can run a custom report at any time. These assessments include sites reported, the total number of items found, how many were accessed and how often, and how many sensitive information types were found and accessed.

Learn more about data oversharing at Microsoft Purview data security and compliance protections for generative AI apps.

5.2.2.3 Microsoft Purview Information Protection

Microsoft Purview Information Protection provides a comprehensive view of the potential risks to an organization's sensitive data. It allows users to automatically discover, label, and protect data based on its sensitivity label across agents using SharePoint grounding data, so they can more effectively manage and reduce overall risk. For agents using SharePoint as their grounding data, Microsoft Information Protection honors view/extract usage rights for any content encrypted, cites sensitivity labels in agent responses, has the sensitivity labels for files referenced in prompts and responses, and labels conversations with the most restrictive sensitivity label.

Learn more about Microsoft Purview Information Protection at Microsoft Purview Information Protection.

5.2.2.4 Insider risk management

Data doesn't move itself—people move and access data. Insider risks are the leading cause of data breaches. In addition to data and permission controls that help address data oversharing or leakage, security teams also need ways to detect users' risky activities in AI applications that could potentially lead to data security incidents. Risky AI usage indicators, policy templates, and analytics reports in Microsoft Purview Insider Risk Management capabilities can help security teams with appropriate permissions detect risky activities across agents, such as receiving an unusual

number of agent responses containing sensitive data. Security teams can then effectively detect and respond to these potential incidents to minimize the negative impact.

Learn more about risk assessment at Get started with insider risk management.

5.2.2.5 Communication compliance

Communication Compliance is a tool designed to ensure that AI-driven interactions adhere to established regulatory and code-of-conduct policies. It helps compliance and risk admins detect and investigate harmful and violent content, copyright violations, unauthorized disclosures, etc. Use cases for Communication Compliance include enabling regulatory compliance in industries such as finance and healthcare, improving customer satisfaction by preventing agent interactions that may result in a security or compliance incident, and safeguarding sensitive information in AI interactions.

Learn more about compliance management at Communication compliance.

5.2.2.6 eDiscovery

Agent interactions can be reviewed in eDiscovery for investigations and litigations. Electronic discovery (eDiscovery), refers to the process of identifying, collecting, and producing electronically stored information (ESI) in response to a request for production in a lawsuit or investigation. This capability is crucial as it allows legal teams to uncover and analyze relevant digital interactions that may serve as evidence.

For example, in cases where agents are involved, legal teams can leverage eDiscovery to process, analyze, and review agent interactions to respond to litigation effectively.

Learn more about eDiscovery at Learn about eDiscovery solutions.

5.2.2.7 Audit

Audit logs record interactions with agents to facilitate investigations whenever necessary. These logs capture essential details such as timestamps, user identities, user prompts, and AI responses. By maintaining comprehensive records, organizations can help with transparency, accountability, and compliance with regulatory requirements. Audit logs also enable administrators to detect and respond to security incidents, track usage patterns, and conduct thorough forensic analysis in case of any anomalies or breaches.

Learn more about audit at Learn about auditing solutions in Microsoft Purview.

5.2.2.8 Data Lifecycle Management

Organizations need to manage their records and information lifecycle carefully to meet legal, business, privacy, and regulatory obligations. This involves retaining what is needed and removing what is not needed across the organization. Built-in classification and governance can help, by automatically classifying content and applying the appropriate policies. With Purview's Data Lifecycle Management, security admins can set retention and deletion policies for their agent interactions.

Learn more about audit at Learn about Data Lifecycle Management solutions in Microsoft Purview.

5.2.3 Agent inventory

The Agent Inventory feature in the Microsoft 365 admin center (MAC) provides administrators with a centralized view of all Copilot agents integrated into the organization's Microsoft 365 environment. It enables streamlined management and governance of agents, ensuring compliance with security and privacy standards.

When an agent is submitted for approval, uploaded by an administrator, or referenced from the public store, its metadata is displayed in the Agent Inventory detail view. This includes:

- Agent capabilities
- Data sources
- Custom actions
- Compliance indicators

Administrators can use this information to make informed decisions about allowing or blocking agents based on functionality and security posture. The Shared Agents page and search tools make it easy to locate agents across departments, evaluate usage, and take corrective actions where necessary.

Blocking or removing non-compliant agents is a critical step in maintaining a secure and efficient IT environment. Continuous monitoring through Agent Inventory helps safeguard sensitive data and uphold organizational standards.

Learn more about Integrated Apps at Get started with Integrated apps - Microsoft 365 admin.

Learn more about managing agents at Manage agents for Microsoft 365 Copilot in Integrated Apps - Microsoft 365 admin.

5.2.3.1 Inventory management

All agents are treated as apps in the system, allowing administrators to have a centralized inventory. The Integrated Apps page of the Microsoft 365 admin center (MAC) lists all apps as well as all agents (SharePoint agents[2] will be added in the near future) available in the tenant, enabling actions like blocking agents or assigning them to specific users. Administrators maintain full app management capabilities for agents, ensuring a catalog of allowed agents and the ability to intervene when necessary.

The available reports provide an overview of the agents that the organization has developed using various tools such as Copilot Studio, Agent Builder, and Teams Toolkit. This comprehensive view allows administrators to see all the agents created under the organization's umbrella.

Learn more about inventory management at Get started with Integrated apps - Microsoft 365 admin.

Learn more about managing agents in Integrated Apps at Manage agents for Microsoft 365 Copilot in Integrated Apps - Microsoft 365 admin.

5.2.3.2 Publisher attested and Microsoft 365 certification

Applications available in the public store must have a publisher attestation and be certified by Microsoft 365.

Publisher attested means that the publisher of the agent has self-attested to the authenticity, security, and reliability of their application. This involves the publisher signing a legally binding document asserting that their application meets certain industry standards and recommended practices for development and testing. This attestation provides a level of assurance to users and administrators that the application has been developed with care and attention to security protocols.

The Microsoft 365 certification, on the other hand, is an additional layer of validation performed by Microsoft itself. Applications that achieve this certification have undergone a rigorous evaluation process to ensure they meet Microsoft's stringent security, compliance, and performance requirements. This certification process

[2] Refer to *Section 3.2.2 SharePoint Online content permissions* for managing SharePoint agents.

involves a thorough assessment of the application's security features, data handling practices, and overall functionality within the Microsoft 365 ecosystem.

Together, these two layers of validation (Publisher Attested and Microsoft 365 certification) help build trust and confidence among users and administrators. They ensure that only high-quality, secure, and reliable applications are deployed within the organization's digital environment, reducing the risk of security breaches and enhancing overall user experience and productivity.

Learn more about Microsoft 365 certification at Microsoft 365 App Compliance Program overview.

5.2.4 Agent sharing

Zone 3 sharing supports broad organizational access to high-value, high-risk, or business-critical agents. Sharing is tightly controlled and typically managed exclusively by central IT or the Center of Excellence (CoE). Access may be granted via:

- **Large-scale Microsoft Entra security groups** (e.g., All Employees, Sales Organization),
- **Role-based access groups**, or
- **Restricted distribution lists** for phased rollouts.

Editor access is rarely permitted in production. Most zone 3 environments allow **viewer/run-only access**, ensuring that only validated owners and maintainers can update the agent. All sharing is backed by environment group rules enforcing connector management policies, identity constraints, and security baselines appropriate for enterprise-wide agents.

Publishing in zone 3 follows formal change-controlled processes. Agents may be:

- Published to the **Microsoft 365 App Catalog**
- Distributed via Teams with targeted app permission policies
- Rolled out through **phased releases** (pilot → department → tenant), or
- Integrated into enterprise Copilot experiences.

Enterprise agents must pass any required security, compliance, or responsible AI reviews before sharing or publishing. Pipelines and solution-based ALM are commonly used to ensure consistent deployment between Dev, Test, and Prod environments.

These controls ensure that enterprise agents operate safely at scale, are only accessed by authenticated users with legitimate business need, and remain protected by the organization's highest governance, monitoring, and operational standards.

Learn more about agent sharing at Share and manage agents.

5.3 Agent reporting

5.3.1 Microsoft admin center

Refer to the zone 2 section 4.4.1 Microsoft admin center Integrated Apps.

5.3.2 Microsoft Purview

Refer to the zone 2 section 4.4.2 Microsoft Purview.

5.3.3 Copilot Studio admin controls

Refer to the zone 2 section 4.4.3 Copilot Studio admin .

5.3.4 Microsoft Sentinel

Microsoft Sentinel is a scalable, cloud-native security information and event management (SIEM) platform that delivers an intelligent and comprehensive solution for SIEM and security orchestration, automation, and response (SOAR). Microsoft Sentinel provides cyberthreat detection, investigation, response, and proactive hunting, with a bird's-eye view across your enterprise.

Microsoft Sentinel also allows administrators to keep a close watch on all agent activities by providing real-time monitoring and alerting capabilities. The platform offers comprehensive insights into potential security threats, suspicious activities, and compliance issues. By setting up customized alerts, administrators are promptly notified of any critical events, enabling them to take immediate action to mitigate risks. Additionally, Microsoft Sentinel's robust reporting tools analyze historical data to identify patterns and improve overall security posture over time.

Learn more about Microsoft Sentinel at What is Microsoft Sentinel?

Learn more about configuring Microsoft Sentinel at Configure Microsoft Sentinel content.

6 Lifecycle governance for Copilot Agents

The lifecycle for Microsoft Copilot Studio agents, including those built with Agent Builder in Microsoft 365 Copilot, follows the stages of Creation, Deployment, and Monitoring and Ongoing Governance.

Each stage is governed through a zoned model using zone 1, zone 2, and zone 3 environments, each configured with specific connector controls, identity boundaries, and data protection requirements.

Detailed procedures for creating, deploying, and managing Copilot Studio agents are provided in the Step-by-Step Admin Guide for Copilot Studio Agents; this whitepaper focuses on the governance concepts.

6.1 Creation: building safely within the right zone

Creation begins by selecting the correct environment (zone) and ensuring platform guardrails are in place.

6.1.1 Environment types

Zone 1 – Personal productivity

- Personal developer environments intended for individual experimentation.
- Connector access is limited by environment-group controls, and sharing is severely restricted.

Zone 2 – Team collaboration

- Managed environments for departmental agents shared with a defined group.
- Makers are scoped via Microsoft Entra groups, and connector availability is centrally governed.

Zone 3 – Enterprise Managed

- Tightly controlled environments for mission-critical or organization-wide agents, owned by central IT.

6.1.2 Baseline governance controls during Creation

Regardless of zone:

Connector security (PPAC)
Enforced by Connector Policies and Advanced Connector Protection (ACPB in PPAC).

These define:

- which connectors are allowed or blocked
- which connectors are classified as high-risk
- whether external or unmanaged connectors may be used
- cross-environment restrictions
- Zone 3 environments apply the strictest ACP settings

Data security (Purview)
Applies to the data an agent can consume—not the connectors.

Purview governs:

- sensitivity labels
- Microsoft 365 DLP
- content inspection and policy-based restrictions
- audit and compliance controls

Identity and authentication

- Zone 1 agents run under the maker's identity
- Zone 2 and 3 agents should use service principals or managed identities configured with least privilege

Prompts and guardrails
Instructions must include scope boundaries and constraints to prevent cross-domain access or data exfiltration.

Design review

- Zone 1: automated guardrails only
- Zone 2: lightweight CoE review

- Zone 3: formal review across security, privacy, and Responsible AI

Where required by internal policy, especially in zone 2 and zone 3, agents go through lightweight or formal design review (security, privacy, responsible AI) before being cleared for broader use.

6.2 Deployment: controlled promotion and scoped access

Deployment determines how agents are shared and how they move through environments.

Zone 1 – Personal productivity

- Sharing disabled or limited to a very small test group
- If broader demand emerges, the agent should be rebuilt or promoted into zone 2, not scaled from a personal environment.

Zone 2 – Team collaboration

- Shared with a specific team or Microsoft Entra group (for example, a private Teams team or departmental group).
- Agents are packaged into solutions and moved with Power Platform ALM or pipelines, with connector re-binding and environment approvals.

Zone 3 – Enterprise Managed

- Deployment is owned by central IT
- May include catalog publishing, pilot groups, and organization-wide rollout, combined with app permission policies for scoped access.

Across all zones, deployment includes

- validating that connector policies (PPAC) and Purview protections behave as expected
- ensuring only intended users can invoke the agent
- registering the agent in organizational inventory with clear ownership, zone classification, and purpose
- communicating boundaries and acceptable use to end-users

6.3 Monitoring and ongoing governance

Once deployed, agents require continuous oversight for security, accuracy, and compliance.

Key governance practices include:

- **Platform monitoring:** Usage, adoption, and environment-level reporting via Power Platform admin center (and Copilot Hub where available).
- **Purview audit and compliance:** Audit logs for conversational activity, connector usage, and identity actions. Purview DLP and sensitivity labels continue to govern data-level behavior.
- **SIEM integration (zone 3):** Enterprise agents may integrate with Microsoft Sentinel for anomaly detection, data exfiltration monitoring, and correlation across systems.
- **User feedback:** Critical for identifying hallucinations, bias, or confusing behavior.
- **Policy updates and prompt adjustments:** Governance teams refine connector policies (PPAC), Purview rules, knowledge sources, and prompt guardrails over time.
- **Incident response:** Agents must have a documented containment approach—disable the agent, revoke connections, investigate via Purview and platform logs, remediate, and update governance.
- **Zone re-evaluation:** Agents may be promoted to a higher zone as adoption grows or data sensitivity increases or retired if usage drops.

6.3.1 Governance outcomes

A lifecycle model grounded in correct connector security (PPAC) and data security (Purview) delivers:

- **Transparency:** Agents are registered, documented, and traceable.
- **Consistency:** Connector policies, ACP, Purview rules, and ALM enforce uniform standards.
- **Resilience:** Monitoring, audit, and incident playbooks reduce risk.

- **Value at scale:** Adoption data and feedback guide investment into agents that deliver measurable benefit.

This approach aligns with Microsoft's recommended governance model and the emerging cross-product ecosystem of Copilot, Power Platform, and Purview.

6.4 Scale via distribution

The Microsoft 365 Copilot Agent Store is a primary discovery and access experience for agents within the Microsoft 365 ecosystem. It provides users with a consistent way to find and access agents, while giving IT administrators the administrative controls needed to manage what is available and to whom.

Agents in the store fall into five categories, each with its own distribution path:

1. Microsoft-built agents are enabled by default for all eligible with Microsoft 365 Copilot licenses and do not require additional admin action to be discoverable. From the Integrated Apps page in the Microsoft 365 admin center (MAC), administrators can block an agent tenant-wide, scope availability to specific user groups, or restore a previously blocked agent. Users may also request access to an agent directly from the store, with IT receiving and acting on those requests.

2. Partner-built agents are submitted via Microsoft Partner Center and once approved by Microsoft, appear in the Agent Registry for IT review. Unlike Microsoft-built agents, they are not automatically available, administrator action is required before they surface in the store. Admins can deploy to all users or specific groups, and block or remove a partner agent at any time. Once enabled, they appear in the "More agents" collection.

3. Custom agents built with Copilot Studio follow an explicit publish-and-approve workflow before appearing in the "Built by your organization" collection. The maker submits the agent from Copilot Studio, at which point it enters a pending state, visible to admins in the MAC under Agents > All > Requests, but not yet visible to end users. Admins review and approve or reject the submission. Approved agents are deployed to assigned users or groups; rejected agents are returned to the maker for revision and

resubmission. Deployed agents can be blocked or removed at any time.

4. Agents built with the Microsoft 365 Agents Toolkit or Agents SDK follow the same governance workflow as Copilot Studio agents. Developers publish via the Teams Developer Portal, the agent enters a pending state for admin review, and after approval it becomes available in the "Built by your organization" collection. Admins retain full control to block or remove at any time.

5. Agent 365 SDK–enabled agents allow enabled agents, regardless of how or where it was built, to integrate with Microsoft 365 Copilot. These agents appear in the "Agents for your team" collection and can be invoked inline within Copilot chat. Admins can scope availability to specific users or groups and block or remove these agents at any time.

Across all categories, the Agent Store serves as the governed entry point that helps organizations scale agent adoption safely, balancing accessibility and innovation with the security and oversight required at each zone of governance.

Learn more about the Agent Store at: https://learn.microsoft.com/en-us/microsoft-365/copilot/copilot-agent-store

6.5 Observability

Agent 365 is the unified control plane for agents that enables IT and security to observe, govern, and secure agents across the organization with proper oversight and guardrails. Agent 365 helps customers to move from agent experimentation to enterprise-scale operations.

With Agent Registry, IT teams can get an inventory of agents in their organization, including agents built with Microsoft AI platforms, ecosystem partner agents, and agents registered via APIs, while Agent behavior and performance observability provides detailed reports about agent performance, adoption and usage metrics, an agent map, and activity details.

7 Getting started

To start empowering employees to create and manage Microsoft 365 Copilot agents securely at scale, it is recommended that a three-phase approach is taken:

1. Form an "agent adoption champion" team within IT and try out Agent Builder.
2. Train employees to build agents with Agent Builder.
3. Provide selected employees with access to Agent Builder and set up pay-go meters.

Figure 7 also provides example use cases for agents that could be created to get started.

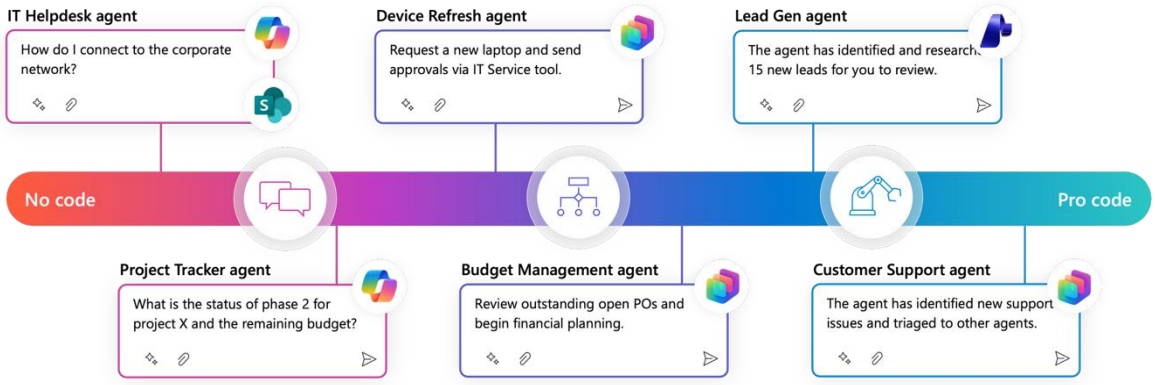


Figure 7 Example agent use cases to get started

Phase I – Form a team

Form an "agent adoption champion" team within IT and try out Agent Builder.

Step 1: Create an "agent adoption champion" team

Select a group of early adopters (your "champion team") to explore building agents using Agent Builder in Microsoft 365 Copilot. Provide them with the required Copilot licenses to test features, develop recommended practices and identify use cases.

Step 2: Provide the right controls to the right audience

Use the Microsoft 365 admin center to assign Copilot extensibility rights exclusively to your selected champion team. By doing so, you can empower a trusted group to

explore advanced Copilot capabilities while ensuring your environment remains secure and compliant.

Step 3: Empower your champion team to build an agent

Empower your champion team to create the first org-wide agent using Agent Builder. This process will both validate recommended practices before broader deployment and provide employees with an initial on-demand resource.

Learn more about building agents with Build agents with Copilot Studio.

Phase II – Train employees

Train employees to build agents with Agent Builder.

Step 4: Gradually enable web-grounded agent development

Provide structured training and recommended practices on Copilot Chat fundamentals and safe, web-grounded agent building to each department in your organization. Then, grant development permissions through Microsoft 365 admin center security groups—ensuring each team is prepared before rolling out this capability.

Step 5: Enable proof-of-concept agent consumption

Use the Copilot Studio admin controls to grant Copilot Chat users access to your initial proof-of-concept agent. This allows IT to gather insights on agent performance and refine capabilities before a broader organizational rollout.

Step 6: Establish a Center of Excellence

Establish a Center of Excellence (CoE) to govern all agents built across the organization. The CoE, led by your champion team, will define standards, approve new agents, and ensure each department follows recommended practices for secure and effective development.

Phase III – Deploy and engage

Provide select employees with access to Agent Builder and set up pay-go meters.

Step 7: Identify and train departmental agent creators

Designate key individuals to develop agents that can access critical work data. Provide them with specialized training before migrating them from Copilot Chat to Microsoft 365 Copilot licenses, equipping them with the tools they need to securely and efficiently build robust, department-specific solutions.

Step 8: Set up pay-go meters and control agent sharing

Configure a dedicated pay-go meter in Copilot Studio for each department. Allow approved agent creators to build solutions using their department's meter and enable limit sharing with the Copilot Studio admin controls to prevent agents from being overshared across the organization.

Step 9: Govern org-wide agent sharing

Invite employees to share the agents they create so the Center of Excellence (CoE) can evaluate each one and decide whether it should remain within a specific group or be accessible across the entire tenant. If agents fail to meet CoE standards, IT can block them at the tenant level from the Microsoft 365 admin center. Once the agents meet the standards, IT can unblock them.

Step 10: Manage spend and monitor usage

IT admins can monitor and manage agent spend with the Copilot Studio admin controls. For example, IT admins can turn on consumption alerts for agents by navigating to 'Copilot Studio', selecting 'Manage Capacity' and setting up alerts based on desired usage thresholds.

8 Conclusion

This whitepaper provides a detailed framework for securing and governing agents within Microsoft 365 environments. By adhering to the specified governance principles and implementing comprehensive security and management controls, organizations can effectively manage their agents while ensuring compliance and operational efficiency. The descriptions of governance zones offer IT practitioners a clear path to progressively enhance their agent management strategies. Ultimately, this document aims to equip IT departments with the expertise needed to navigate the complexities of agent governance, thereby fostering a secure and well-governed Microsoft 365 ecosystem.

The whitepaper delves into various aspects of agent governance, beginning with an overview of the foundational principles that underlie robust security practices. These principles emphasize the importance of maintaining strict access controls, continuous monitoring, and regular updates to ensure that all agents operate within predefined security parameters. Additionally, the paper outlines the roles and responsibilities of different stakeholders in the governance process, highlighting the collaboration required between IT, compliance, and business units.

Furthermore, the whitepaper explores advanced security measures which are crucial for safeguarding sensitive data and preventing unauthorized access. It addresses the challenges posed by emerging threats and provides strategies for mitigating risks through proactive and reactive measures while also focusing on observability for IT teams.

In summary, this whitepaper is a comprehensive guide designed to help organizations establish a secure and efficient framework for managing agents within Microsoft 365. By following the recommendations and recommended practices outlined in the document, IT departments can achieve higher levels of security, compliance, and operational effectiveness in their digital ecosystems.