Microsoft

Administering and Governing Agents

1 Introduction

1.1 Purpose

This document aims to provide a comprehensive overview of the strategies and tools necessary for securing and governing 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. It covers the role of Microsoft 365 agents, Agent Builder agents, and Copilot Studio agents, explaining how each type can be utilized to enhance operational efficiency while maintaining stringent security protocols.

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.

1.2 Scope

This whitepaper focuses on governing agents within the Microsoft 365 ecosystem. This includes agent built with:

- **SharePoint** – Use SharePoint to build agents based on content stored in SharePoint sites and libraries. Learn more at Get started with SharePoint agents.
- **Copilot Studio Agent Builder** – Use Copilot Studio Agent Builder tool directly within M365 Copilot to create conversational templates tailored to specific

tasks or business needs. Learn more at Use Copilot Studio Agent Builder to Build Agents.

- **Full Copilot Studio** – Use triggers, advanced logic, and connections to other Microsoft services or third-party platforms to create agents with full Copilot Studio. Learn more at Overview - Microsoft Copilot Studio.
- **Pro Developer Tools** – Use tools and services like Team Toolkit and Azure AI Fabric 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

When providing governance to agents, organizations should consider: the user audiences that create agents; the tools that they use; and the capabilities that the tools provide.

There are three groups of user audiences that create agents:

- **End Users** – These are individuals within an organization who use intuitive tools—such as SharePoint and Agent Builder—to create solutions that support their daily tasks. These tools are designed to enhance productivity without requiring deep technical expertise. Everything they interact with operates within the boundaries of their existing permission structure, ensuring secure and appropriate access.
- **Makers** – Individuals within an organization who utilize more complex tools and have a deeper understanding of technology and governance. These users predominantly work in Copilot Studio to create more advanced solutions but may also use Agent Builder—often taking things further than End Users by incorporating more advanced features. Makers are distinct from End Users in that they can create autonomous agents that use triggers and handle more complex scenarios.
- **Developers** – Individuals within an organization who engage in more advanced development tasks and have a deep understanding of technology and governance. They are the most advanced user group, possessing the skills and tools to create and manage sophisticated solutions while adhering to governance and control frameworks. Their agents are managed and made available through centralized IT catalogs.

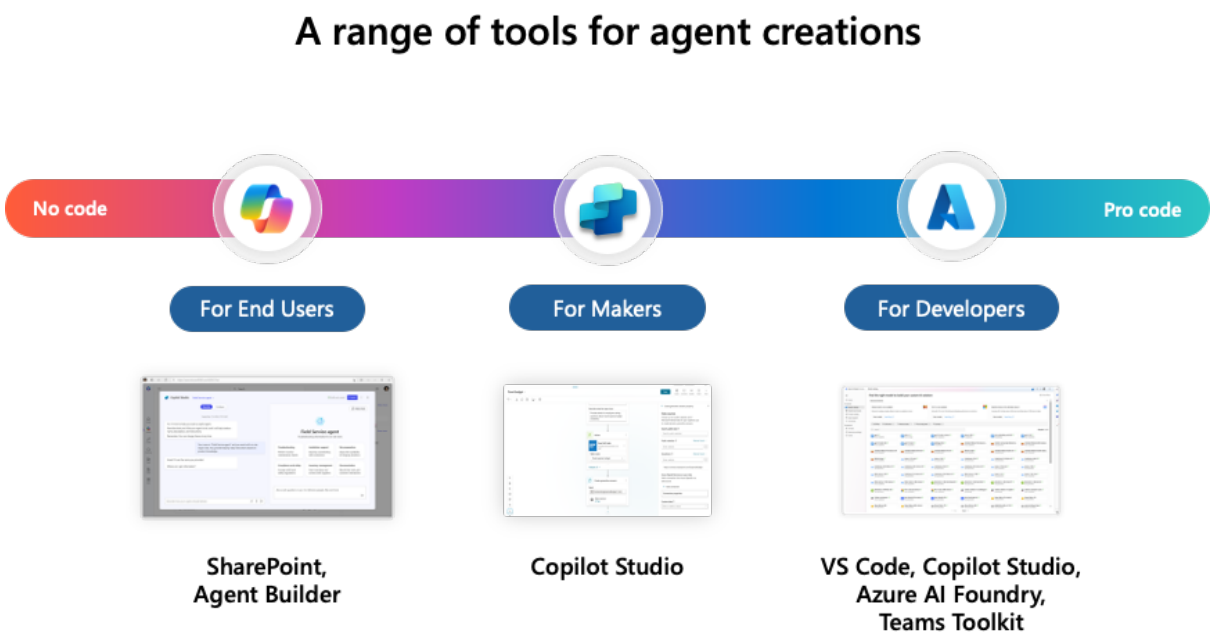

Figure 1 A range of tools for agent creations

These user audiences require different levels of governance based on the complexity of the types of agents they are creating. For example, End User agents leveraging the governance already in place for SharePoint Online, versus IT Catalog agents being centrally managed by IT through the Microsoft 365 Admin Center (MAC). Regardless of the type of agent, they can be governed using a spectrum of controls:

1. **Tool controls** – Govern the deployment and functionality of the agents themselves, ensuring they operate within the defined parameters set by administrators. These controls allow for precise management of the technology infrastructure within which the agents operate, ensuring security, efficiency, and compliance with regulatory standards. Tools controls are primarily managed using the Microsoft 365 Admin Center (MAC) and Power Platform Admin Center (PPAC).
2. **Content controls** – Govern the information processed and generated by agents. They ensure that the data handled by the agents adheres to organizational guidelines and privacy regulations, safeguarding sensitive information and maintaining data integrity. Content controls are primarily managed using the Microsoft 365 Admin Center (MAC), Power Platform

Admin Center (PPAC), Microsoft Purview, and SharePoint Advanced Management.

3. **Agent Management** – Governs controlling and monitoring agents' activities. This includes: allowing for a staged rollout to ensure controlled deployment; reviewing detailed usage reports that provide metrics on user engagement and per-agent activity; aiding adoption tracking; identifying underused agents; supporting chargeback or forecasting; and lifecycle management including deletion of agents. Agent Management is primarily managed using the Microsoft 365 Admin Center (MAC).

The spectrum of user audiences and the types of agents and controls (as per Figure 2 Spectrum of Agents and Controls) allows organizations to maintain a robust, secure, and compliant digital environment while leveraging the capabilities of advanced autonomous agents.

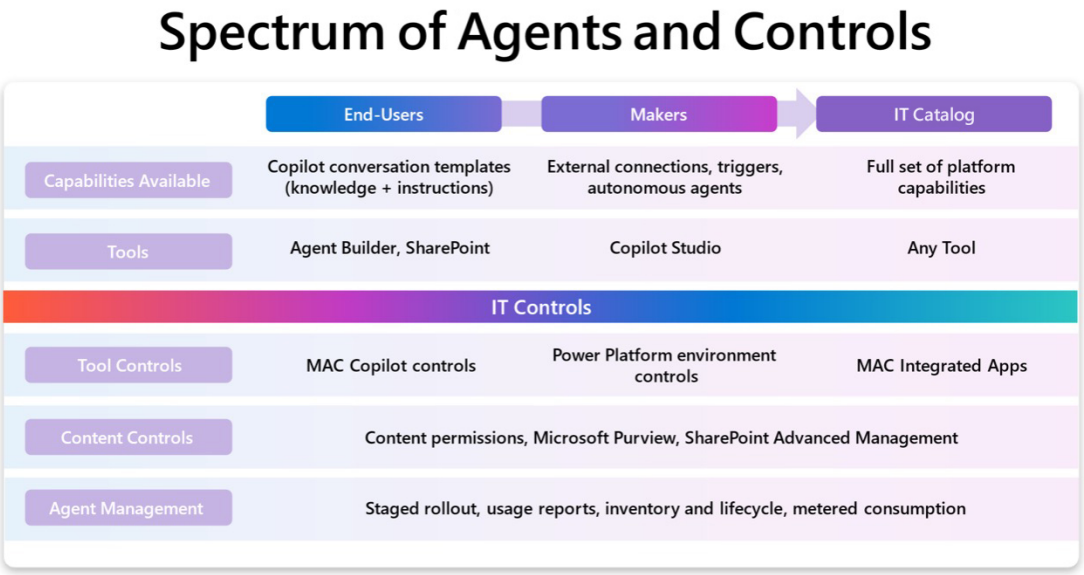

Figure 2 Spectrum of Agents and Controls

3 Microsoft 365 Copilot

Microsoft 365 Copilot is an AI-powered productivity assistant that integrates with the Microsoft Graph and Microsoft 365 apps to enhance creativity, efficiency, and collaboration. It brings intelligent assistance into everyday tools like Word, Excel, Outlook, Teams, SharePoint, and more, helping users write, analyze, summarize, and collaborate with ease. For IT administrators, the Microsoft 365 Admin Center includes controls to manage data access, security, compliance, and agent usage, supporting the safe and scalable deployment of AI across the enterprise.

Learn more about Microsoft 365 Copilot at What is Microsoft 365 Copilot?

3.1 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 Copilot Control System

The Copilot Control System is a system of integrated controls and capabilities for Copilot and agents. Several of its capabilities reside within the Microsoft 365 Admin Center (MAC), where administrators can find comprehensive oversight and management capabilities that enable them to monitor and regulate system activities with precision. By leveraging usage and inventory analytics, the Copilot Control System empowers admins to identify usage trends and agent adoption in their organization.

Furthermore, these controls streamline operational workflows, reducing administrative burden and ensuring that resources are utilized efficiently. Automated processes within the Copilot Control System facilitate rapid response to incidents, minimizing downtime and maintaining system integrity.

Learn more about the Copilot Control System at Copilot Control System.

Learn more about the Copilot Control System in MAC at Microsoft 365 admin center scenarios that configure Copilot.

3.3 Integrated Apps

The Integrated Apps administration functionality within the Microsoft 365 Admin Center (MAC) is designed to work seamlessly within the Microsoft 365 ecosystem, providing users with a streamlined and cohesive experience. It is also used to primarily manage the IT Catalog agents[1].

Whenever an agent is submitted for administration approval, uploaded by an administrator, or referenced from the public store, all metadata about the agent's definition is provided on the MAC Integrated Apps app detail tab. This equips

[1] Note that SharePoint agents are not available to be managed in the Integrated Apps administration functionality. Refer to 3.4 SharePoint Online Content Permissions for managing SharePoint agents.

administrators with comprehensive information about the agent, including its capabilities, data sources, and the custom actions that the agent can invoke. Furthermore, administrators can review the agent's compliance with security standards and privacy policies, ensuring that only trustworthy and efficient agents are integrated into the organization's ecosystem. This detailed metadata helps administrators make an informed decision about allowing or blocking the agent based on its functionality and security posture.

Administrators can also utilize the shared agents page or the search functionality to look up specific agents by name, making it straightforward to identify and manage them within the Integrated Apps section. This feature is particularly useful for tracking down shared agents that may have been distributed across various departments or user groups. By locating these agents swiftly, administrators can evaluate their usage and compliance and take necessary actions to block any shared agents that do not meet the organization's security or operational standards. This management approach ensures that only approved and safe agents are deployed within the environment, safeguarding the integrity of the organization's digital ecosystem.

Blocking shared agents is also a crucial step in maintaining a secure and efficient IT infrastructure. Shared agents can sometimes be overlooked, leading to potential vulnerabilities or unauthorized access. By thoroughly reviewing and managing these agents, administrators can ensure that all active agents comply with company policies and contribute to a robust cybersecurity posture. This continuous monitoring and management process is essential for protecting sensitive data and maintaining operational integrity.

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.

3.3.1 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 (excluding SharePoint agents[2]) 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 report provides 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.

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

[2] Refer to 3.4 SharePoint Online Content Permissions for managing SharePoint agents.

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.

3.4 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.

For SharePoint agents, this means that by default agents can only access content or update data that the current user already has access to.

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.

3.5 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 also includes tools like site sharing restrictions and block download policies to prevent oversharing of confidential information. It can also 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.

3.6 Usage Reports

The Microsoft 365 Copilot page in the Microsoft 365 Admin Center (MAC) - Reports - Usage provides reporting on agents and usage. Administrators can monitor agents and those shared in Microsoft 365, and can also track agent usage frequency, duration, distribution across platforms, thus gaining insights for better resource allocation and efficiency.

For instance, by analyzing user engagement metrics, administrators can determine which agents are most effective and widely adopted within the organization. This can help in making data-driven decisions about which agents to further develop or promote into the IT Catalog. Additionally, tracking the duration and frequency of agent usage allows administrators to understand how often employees rely on these

tools, ensuring that they are providing value and justifying the resources invested in their development.

Through the MAC, administrators can also observe the distribution of agent usage across various platforms such as desktop, web, and mobile. This insight is crucial for identifying platform-specific performance issues or user preferences, enabling targeted improvements and support.

Overall, the comprehensive reporting and monitoring capabilities provided by the MAC empower administrators to optimize agent deployment, enhance productivity, and ensure that resources are effectively aligned with organizational needs.

Learn more about usage reports at Microsoft 365 admin center activity reports - Microsoft 365 admin.

Learn more about Copilot usage reports at Microsoft 365 admin center Microsoft 365 Copilot usage - Microsoft 365 admin.

3.6.1 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 help in analyzing 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.

3.7 Cost Management

Administrators can select either pre-paid user licenses or metered billing based on actual usage. This flexibility enables customers to manage adoption and costs effectively.

Pre-paid user licenses allow administrators to purchase a set number of licenses in advance, providing a predictable cost structure and ensuring that all users have access to the necessary resources. On the other hand, metered billing offers a pay-as-you-go model where charges are based on the actual usage of services. This approach can be particularly advantageous for organizations with fluctuating needs or those that want to avoid upfront costs. These two options allow administrators to choose the most suitable billing method to align with their budgetary constraints and usage patterns, ultimately optimizing both resource allocation and financial planning.

3.7.1 Metered Consumption

Metered Consumption is a consumption-based billing model for Copilot Chat agents, enabling pay-per-use access for unlicensed users. Administrators can choose between pay-as-you-go or prepaid plans via the Power Platform Admin Center, allowing cost management and scalable deployment.

Suitable for businesses with varying support needs, this model provides flexibility without yearly commitments. Pay-as-you-go charges based on usage, while prepaid capacity offers predictable costs for clear usage forecasts.

This approach encourages widespread use of Copilot Chat agents, improving customer service, productivity, and resource management without upfront investments. It aligns expenses with usage, supporting strategic implementation.

Learn more about Metered Consumption at [Microsoft 365 Copilot pay-as-you-go overview](#).

3.7.2 Free vs. Premium Functionality

Without a metered consumption plan, users without a Microsoft 365 Copilot license are limited to the free Copilot Chat experience. The free experience includes web-grounded answers and simple Q&A agents, while features requiring access to Microsoft 365 data are blocked. This separation helps organizations pilot Copilot Chat broadly while controlling the rollout of data-connected agents.

4 Copilot Studio

Copilot Studio is a powerful addition to the Microsoft Power Platform family and is a graphical, low-code tool for building agents and agent flows. It integrates directly and inherits from the Power Platform administrative features for security and governance.

Learn more about Copilot Studio at Overview - Microsoft Copilot Studio.

4.1 Power Platform Admin Center

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 makers. Through robust controls, administrators can define and enforce policies that dictate agent behavior, ensuring compliance while maintaining flexibility for innovation. The PPAC provides tools to classify and protect sensitive data through Data Loss Prevention (DLP) policies, safeguarding interactions without hindering productivity. Additionally, granular governance features allow administrators to monitor and restrict unauthorized actions, ensuring agents – including chatbots and automated workflows – operate security 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. 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 also a key feature available for governance of agents, specifically within the PPAC, in that administrators can enable a billing plan to monitor consumption of agents and agent messaging.

By implementing these measures, the PPAC 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) - Power Platform.

Learn more about Power Platform data storage and governance at Data storage and governance in Power Platform - Power Platform.

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

Learn more about using the Power Platform Admin Center at Overview of the Power Platform admin center.

4.2 Power Platform Environment Controls

Power Platform 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, data loss prevention 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.

Power Platform environment controls are inherited by Copilot Studio and are required to maintain secure and efficient operations within all the Power Platform tools. 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, Power Platform environment controls help prevent unauthorized activities and ensure compliance

with organizational standards. These controls extend to each component in the Power Platform product stack, including 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, makers, 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, Data Loss Prevention (DLP), 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 Power Platform environment controls at Managed Environments overview - Power Platform.

4.2.1 Authoring agents

Admins can effectively provide the toolset for makers to create and author agents by creating environments and environment scopes, setting environment roles, renaming the Default environment and monitoring the environment history from within the Power Platform Admin Center (PPAC). This can include a welcome message that informs makers and Copilot Studio makers of the policies and practices of the environment that have been shared with them.

Organizations should ensure that each Copilot Studio maker 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[3].

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 - Power Platform.

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

4.2.2 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 Power Platform Admin Center (PPAC) governs both makers and the agents they create within Power Platform and Copilot Studio – ensuring innovation stays within guardrails. Administrators use security groups and rules to control who can modify artifacts within environments (including the environments themselves), preventing unauthorized changes to Copilot Studio agents, chatbots, or workflows, while makers (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 maker can override organizational standards. Ensuring makers can build agents in personal environments is also a way to ensure other makers 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 Power Platform Environments have been created, administrators can use the PPAC Deployment page to create and establish Pipelines.

[3] Environment management and routing is covered in Section 4.2.2 Environment Management.

Once a Pipeline(s) has been established, Makers 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 Power Platform environments overview.

Learn more about environment routing at Power Platform environment routing.

Learn more about Pipelines at Overview of pipelines in Power Platform.

4.2.3 Agent Sharing

Share agents only with those involved directly to ensure information security within collaborative environments. Co-authors and chat end users require access to contribute effectively and engage in interactions. For instance, co-authors might need to share drafts, provide feedback, and incorporate changes suggested by their peers. End users must have ongoing access to be able to continue using the agent. Sharing limits should be established ensuring that Editor and Viewer permissions are restricted, or the number of times an agent can be shared.

Limiting access is a key risk management practice that helps preserve data integrity, prevents unauthorized changes, and reduces the likelihood of sensitive information being exposed or distributed inappropriately.

Learn more about agent sharing at Publish and Manage Copilot Studio Agent Builder 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.

4.2.4 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.

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.

4.2.5 Power Platform Data Policies

Data Loss Prevention (DLP) policies let users govern how agents connect to and interact with data and services—both within and outside the organization. Administrators can configure Copilot Studio and Power Platform DLP policies in the Power Platform Admin Center (PPAC).

DLP policies in PPAC can be configured to govern the use and availability of Copilot Studio features and agent capabilities. This allows administrators to control what agents can access and how they interact with enterprise resources. The following features and capabilities are included:

- Maker and user authentication
- Knowledge sources
- Actions, connectors, and skills
- HTTP requests
- Publication to channels
- Application Insights (AppInsights)
- Triggers

Copilot Studio connectors can be classified within a DLP policy into the following data groups, visible in the Power Platform admin center when configuring DLP policies:

- Business
- Non-business
- Blocked

By configuring connectors in DLP policies, administrators can protect the organization's data from malicious or unintentional data exfiltration by agent makers.

Learn more about Copilot Studio security and governance at Key Concepts - Copilot Studio security and governance

Learn more about managing data policies at Manage data policies - Power Platform.

5 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 Microsoft 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 about Microsoft Purview at Microsoft Purview.

5.1 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 & data security and compliance protections for Microsoft 365 Copilot and Microsoft 365 Copilot Chat.

5.2 Data Loss Prevention

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.

5.3 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.

5.4 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.

5.5 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 template, and analytics report in Microsoft Purview Insider Risk Management capabilities can help security teams with appropriate permissions to 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.

5.6 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.

5.7 eDiscovery

Agent interactions can be reviewed in eDiscovery for investigations and litigations. eDiscovery, or electronic discovery, 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.

5.8 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.

5.9 Data Lifecycle Management

Organizations need to manage their records and information lifecycle carefully to meet legal, business, privacy, and regulatory obligations. That involves keeping what they need and getting rid of what they don't need broadly 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.

6 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 3 also provides example use cases for agents that could be created to get started.

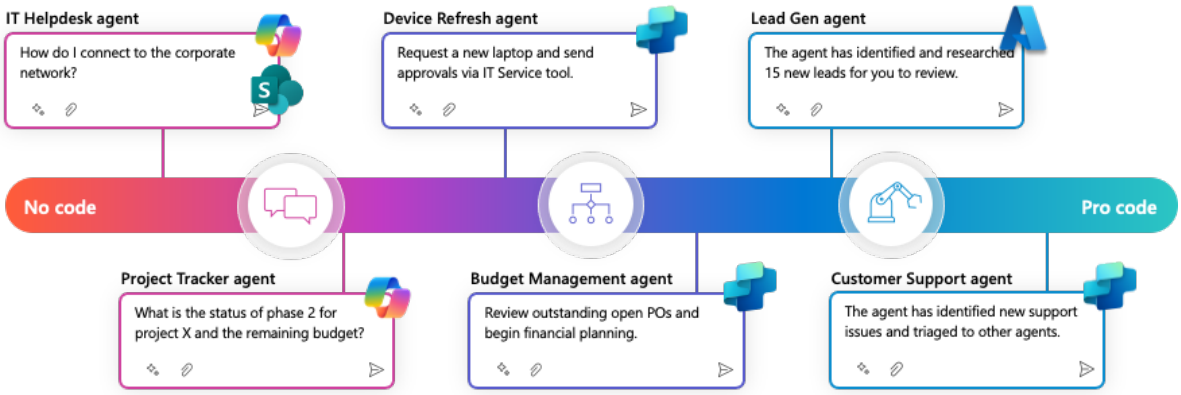

Figure 3 Example agent use cases to get started

6.1 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 Microsoft Copilot Studio Agent Builder in Microsoft 365 Copilot. Provide them with the required Copilot licenses to test features, develop best 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 best practices before broader deployment and provide employees with an initial on-demand resource.

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

6.2 Phase II – Train employees

Train employees to build agents with Agent Builder.

Step 4: Gradually enable web-grounded agent development

Provide structured training and best 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 Power Platform Admin Center 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 best practices for secure and effective development.

6.3 Phase III – Deploy and engage

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

Step 7: Identify and train departmental agent makers

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 makers to build solutions using their department's meter and enable limit sharing in the Power Platform Admin Center 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 in the Power Platform Admin Center. 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.

7 Conclusion

Agents are transforming work by automating repetitive tasks, allowing humans to focus on strategic activities. They standardize operations, reduce errors, and adapt to changes, integrating with various systems. Agents collect and analyze real-time data for better decision-making and proactive problem-solving, helping organizations stay competitive. They also enhance team collaboration by streamlining communication and information sharing, enabling effective teamwork and a cohesive work environment.

In this changing world, administrators have a pivotal role in fostering innovation (and hence agents) through governance frameworks. It is essential for IT professionals and decision-makers to understand that governance should not stifle creativity but rather provide a structured pathway that enables the seamless integration of new technologies.

By demonstrating the strategic importance of governance frameworks, administrators can champion the development of agents that enhance productivity while maintaining high standards of security and compliance. This dual focus on innovation and governance ensures that new agents are not only effective but also sustainable and aligned with organizational goals.

In conclusion, this whitepaper presents a framework for the management and governance of agents within Microsoft 365 environments. It addresses the key concerns of IT professionals and decision-makers by underscoring the necessity of robust governance strategies, regulatory compliance, and the maintenance of a secure digital ecosystem.