Microsoft

Microsoft IQ and the Agent Data Platform

Build and scale enterprise agents that can understand and act on your business context with Microsoft IQ and Microsoft Agent Factory

Table of Contents

Executive Summary

AI agents are moving beyond simple prompt-and-respond experiences. In production environments, agents plan work, retrieve knowledge, and execute multi-step business processes on behalf of and in collaboration with users and teams. At Microsoft, we designed our agent strategy to help organizations move from experimentation to production with confidence. That strategy is built on a shared intelligence foundation: Microsoft IQ. The components of Microsoft IQ – Work IQ, Fabric IQ, and Foundry IQ – work in harmony across your data estate to make raw data more meaningful and actionable for agents while maintaining protection for your data at every stage. (See how Microsoft is empowering frontier transformation)

This whitepaper is written for IT leaders, architects, and platform teams responsible for deploying agents safely at scale. It will show you how to:

- **Design a layered agent architecture** that evolves from simple work agents to autonomous, multi-agent systems.
- **Unify your data estate** to eliminate silos and complexity, and train agents on data spanning your entire organization.
- **Apply security by design** across identity, data protection, tool access, observability, and governance.
- **Use implementation checklists and stage gates** to move agents through prototype → pilot → production → scale.
- **Learn from real customer examples** that demonstrate how these patterns work in practice.

Our goal is to help you build AI agents that deliver real business value—**without compromising trust, control, or operational rigor**.

Why you need an agent data platform (not just a Model)

As large language models become widely available and easier to integrate, building a basic AI agent is fairly easy. At the same time, differences between one model and another are narrowing.

This is because models are only trained on public data – but what makes your organization unique and special is your trusted, enterprise data. The missing piece to enterprise-ready AI is your IQ. For your agents to be most effective, you need to bring your IQ to the models – the systems that provide trusted context, enforce control, and help ensure AI behaves predictably in real-world business environments.

To create agents at scale that you can trust with authority, agents must be

			
Grounded in **authoritative data** and **shared business context**.	Enforced by **permissions at runtime** (not only at indexing time).	Limited to **governed, auditable actions** when they take action.	Managed as part of a **growing agent** estate (including third-party and custom agents).

For these reasons, we recommend thinking in terms of an **agent data platform**, not just a language model or agent framework. Our approach brings **data and context**: Work IQ, Fabric IQ, and Foundry IQ provide the shared intelligence layer that agents rely on, while unifying your underlying data estate. The result is an architecture that lets you steadily increase agent capability and autonomy over time.

Microsoft IQ: Evolving agents without losing control

As agents mature, their complexity and autonomy increase. We designed Microsoft IQ to map directly to this progression so you can grow an agent's capabilities in a controlled, predictable way:

Work IQ is the intelligence layer that personalizes Copilot and agents to you and your organization's full work context – people, content, and collaboration – Inside Microsoft 365, enabling fast, high-value scenarios in the flow of work. Microsoft Ignite 2025 announcement for Work IQ.

Fabric IQ gives agents unified business context of what's happening across the business right now, how entities are related, what rules govern them, how performance is measured, and how plans and operations are tracking. Introducing Fabric IQ

Foundry IQ enables agents to access, process and act on knowledge from anywhere. With built-in enterprise security and agentic retrieval, agents deliver better results from the right context. What is Foundry IQ?

Microsoft IQ allows you to progress from task-focused assistants to coordinated multi-agent systems, while maintaining enterprise-grade governance at every step.

Work IQ – Build fast, context-rich agents across the flow of work

Work IQ is the intelligence layer that personalizes Copilot to you and your organization, staying with you **across the flow of work**. It powers Microsoft 365 Copilot and agents with a deep understanding of your role, your workflows, and your business, so Copilot reflects the real state of work, not just the words in a prompt.

Work IQ draws on real time data from across the enterprise to track what's in motion, what's been decided, what needs attention, and how best to execute actions. With Work IQ, Copilot moves beyond drafting content to become an active work partner that keeps teams aligned and work flowing smoothly. By capturing enterprise data spanning files, emails, meetings, chats, people and org structure, and connected business systems, Work IQ builds deep understanding from how people and teams operate over time and applies that context directly in daily workflows.

Context captures durable understanding from behavior, not just configuration: derived from real interaction patterns and collaboration, creating implicit understanding of priorities without relying on simply explicit setup. All your data is permission-trimmed, sensitivity label aware, and tenant bounded by default, enforcing Microsoft identity, compliance, and policy controls end to end.

In this section, we will take a closer look at Work IQ. is composed of these tightly integrated layers working together:

Unified Work Data is the secure access to and understanding of both structured and unstructured data from Microsoft 365, Dynamics 365, Power Platform, and other connected business systems (via connectors) that represent work happening across your organization.

Real Time Context is an additional, always-evolving layer of insights that enhance the speed and accuracy of Copilot's response to queries (via memory, semantic index, business understanding). Work IQ helps Copilot learn how people and businesses work – the projects that are important, the frequency of collaboration, who they work with and for, critical workflows, the velocity of communication and much more.

Unified Work Data: Microsoft Graph and Microsoft Dataverse

Rather than treating productivity data and business data as separate silos, Work IQ unifies these work data into real-time context so downstream reasoning can operate across the full scope of work. Microsoft Graph is one key piece of this data layer: It provides access to the unstructured productivity data in Microsoft 365 (SharePoint files, Outlook emails, Teams chats and meetings, user profiles from Entra ID, etc.)

Another key piece is Microsoft Dataverse, which captures structured operational data from line-of-business systems — including Dynamics 365 (Customer Service, Field Service, Finance, and Supply Chain applications) and Power Platform (Power Apps, Power Automate, Power Pages). Dataverse models the entities, relationships, roles, and process state that reflect how a business actually operates. Dataverse enables Copilot and agents to reason over shared business context from Dynamics 365 and Power Apps, including business entities such as accounts, opportunities, cases, inventory, SKUs, price lists, revenue, costs, SLAs, and approvals.

With this access, Copilot and agents will have the ability to reason across both Microsoft 365 productivity data and the business data generated by their systems of record, making it possible for Copilot to answer complex questions like "Help me evaluate how issues raised by my parts supplier in our Teams call last week might impact my inventory and sales in the coming months" and to provide very detailed, specific answer connecting business communications with business data. See how Work IQ connects data from Microsoft 365 with operational data from Dynamics 365 and Power Apps.

But Work IQ's data layer doesn't just include data from Microsoft apps, it is also powered by connectivity to non-Microsoft sources of data (federated data). Microsoft 365 Copilot connectors (200+ available) and Power Platform connectors (1,600+ available) provide built-in, governed connectivity to thousands of third-party enterprise systems beyond Microsoft apps. Data ingested into the customer tenant using Copilot connectors is also included in the semantic index. This enables Copilot and agent to search, reason over, and act on more of your unified enterprise content. Here's why this matters: Your critical data from external enterprise systems stay within the tenant boundary. We're bringing intelligence to where your data already is and reasoning across those disjointed systems. We're bringing intelligence to where your data already is and reasoning across those disjointed systems of record.

Create Context: Memory, Semantic Index, Business Understanding

Build Explicit and Implicit memories in Microsoft 365 Copilot and agents

Work IQ gives Copilot the context of your organization: How people and businesses work, the skills they have, the projects that are important, the frequency of collaboration, who they work with and for, critical workflows, the velocity of communication and much more. For example, Copilot can "show me my important emails," based on several factors: Emails from my manager, active customer threads, emails marked "urgent". The key difference:". Copilot remembers what's "important" to you.

Copilot's **memory** is designed to further enhance Copilot's ability to tailor its experience to Copilot users in Copilot Chat and Copilot in the M365 apps. Memory is constructed from a combination of explicit memory and implicit memory.

- **Explicit memory** is provided to Copilot by the user. A user can personalize Copilot by creating "Custom Instructions" – For example a user may manually add an instruction to "Only provide responses to prompts in the active tense". Alternatively, a user can also create "saved memories". For example, prompting Copilot to "remember that I do not like responses in the passive tense" will result in Copilot creating a saved memory "Prefers responses written in active voice; dislikes passive tense". In each case, Memory is explicitly created by a user action.
- **Implicit memory** is provided to Copilot by using the chat history to infer a durable body of insights. As the body of insights grows, Copilot can provide increasingly personal responses and actions.

Beyond chat history, Copilot incorporates **activity** – like workflows – that help increase the fidelity of Copilot's memory. In the coming months, we will start incorporating other activity patterns generated from all of your everyday apps including Teams and Outlook. Stay tuned for additional updates on enhancements to Copilot memory.

Enable semantic search for Microsoft 365 Copilot and agents

Another key component for Work IQ is that it's not only unified work data but also semantically indexing that same data. **Semantic indexing** is a method of organizing and retrieving data by understanding its contextual and conceptual meaning rather than relying solely on exact keyword matches. This is possible with advanced techniques like vectorization to represent data points numerically, enabling the clustering of semantically similar items in a multi-dimensional space.

To enable agents with your trusted, near real time data, semantic indexing is applied automatically to both Microsoft Graph and Microsoft Dataverse. Rather than relying solely on keyword or lexical matching, the semantic index enables Copilot to perform meaning-based retrieval of your data. For example, Dataverse Search is enriched with an intelligent semantic layer that understands the business data schema and adds true business data understanding for agents to operate. AI powered Dataverse Search transforms your trusted productivity and business data by turning raw information into high-value, semantically indexed data to power actionable intelligence. The result: Superior search results.

Learn more about semantic indexing for M365 Copilot

Build business understanding with Dataverse

Unifying and semantically indexing your data are just the starting points. To unlock real intelligence, AI needs business understanding (your **business semantic layer**) — so Copilot and agents can speak your organization's unique business language to answer questions and take action with precision. Dataverse makes this possible out of the box, across three dimensions:

- **Business Ontology** Dataverse stores the entity relationships and metadata that encode how your business ontology concepts connect to one another — accounts to contacts, opportunities to products, work orders to assets. This rich relational structure is automatically understood by M365 Copilot and agents, navigated by both Dataverse Search and the Dataverse MCP Server, so agents can traverse your business graph to answer complex, multi-entity questions without any additional configuration.
- Glossary configuration: Glossary definitions paraphrase the terminology in your Dataverse tables so agents better understand user questions and respond more accurately. This is especially powerful for acronyms, custom field names, and complex business rules that only make sense inside your organization — teaching agents exactly how your business thinks, not how a generic model would guess. For example, when a

marketing specialist asks about "MPF," the agent knows that means "Marketing and Positioning Framework" and referenced the exact document.

- Synonyms configuration: Synonyms map the natural language your users actually speak to the underlying schema elements in your data model, so agents can bridge the gap between how people ask questions and how data is structured. For example, when a field rep asks about "open tickets," the agent knows exactly which table, column, and status values to look for — no data expertise required.

Together, these three layers transform Dataverse from a data store into a true b**usiness intelligence foundation** — giving every AI agent the context it needs to act like a trusted colleague who knows your business, not just a search engine pointing at your data.

In addition, Copilot and agents should return responses not only grounded in our data, but also in our **business processes and preferred outputs**. Work IQ's context layer addresses this by capturing procedural knowledge from existing business workflows as contextual information.

For example: When a sales leader is drafting a proposal, they might normally perform several steps manually in order to get started: (1) Research – Check the customer contact's role on LinkedIn, review recent CRM interactions, and find relevant messaging or promotions; (2) Approvals – Because this is a key account with existing contracts, they would need to consult leadership for approval or guidance; (3) Optimize for reading – Ensure the email or document is concise and mobile-friendly.

These kinds of implicit steps and best practices are part of an organization's procedural knowledge. Work IQ captures this hidden procedural knowledge and dynamically loads these instructions to answer relevant sales questions so that Copilot can perform domain specific tasks at the level of any expert (in this case the sales lead).

Try it out on the Dataverse MCP preview endpoint: aka.ms/dataversemcppreview.

Implementation checklist: Work IQ and agent governance

Plan

- Identify high-impact agent scenarios that would benefit from *work context* (for example, using organizational data from Microsoft Graph or business data from Dataverse).
- Document your existing workflows to pinpoint where agents could assist or automate tasks. Learn more about assessing workflows for Copilot and agents.

Deploy

- Take advantage of native integration points. Early agents can be surfaced directly within **Microsoft 365 Copilot** experiences or in **Dynamics 365 "in-app"** Copilot experiences, staying entirely within familiar interfaces.
- Copilot and agents, over time, learn your preferences through **explicit and implicit memories** to return tailored responses grounded in your context.

Build

- Connect your enterprise data sources using Microsoft 365 Copilot connectors and Power Platform connectors so agents can access both productivity and business data (with proper security).
- Enable **Dataverse Search** (the AI-enhanced semantic search) to ensure agents retrieve the most relevant business data. Learn More
- Build **business understanding** – For instance, define key business entities and actions via Dataverse so agents can reliably interact with them. Learn More

Operate

- Run regular evaluations of your agent's performance (using test queries or telemetry) to observe how well it's answering and where improvements are needed.
- As you develop more agents, reuse the business logic and context you've built across multiple agents; update these templates or knowledge sources in one place rather than redeploying agents individually.

Microsoft Fabric IQ: Turn your data into semantic knowledge ready for AI

As frontier organizations train agents on their enterprise data, it's clear that data quality and context matter far more than sheer data volume. With Work IQ capturing the activities, events, conversations, signals, and workflows currently in motion, **Fabric IQ** focuses on what's happening across the business right now. By defining core business entities, their live state, and the actions associated with them, Fabric IQ connects data to real-world concepts. Agents use this framework to understand current conditions and take the right action. Together, Work IQ and Fabric IQ ensure agents have both the **operational context** of how work gets done and a **factual view** of the business.

Microsoft Fabric is an end-to-end, unified data and intelligence platform that brings together data integration, data analytics, real-time data, databases, business intelligence, and AI tools needed for modern data projects. Within the Fabric platform, **Fabric IQ** provides the shared semantic foundation, or ontology, that gives data consistent business meaning, so agents and people interpret information the same way across the organization.

Before diving deeper into Fabric IQ, it's important to note that building a holistic, AI-ready data foundation usually involves four core steps:

- **Unify your data estate** – Eliminate silos and reduce architectural complexity by bringing data together.
- **Process and harmonize data** – Stream, integrate, and transform historical and real-time data so it's AI-ready
- **Create semantic meaning** – Add business context to data through ontologies and semantic models, so agents understand data the same way your teams do. *(This is where Microsoft IQ becomes crucial.)*
- **Empower AI agents to act** – Leverage that unified, context-rich data to automate workflows, accelerate decisions, and take autonomous actions in your operations

Microsoft Fabric is designed to help your data teams at every part of this journey—helping you unify your data estate, enrich it with meaning, and power a new generation of AI agents that drive true enterprise transformation.

Unifying your data estate with Microsoft OneLake

In any AI journey, the first major step is often the same: Figuring out **where all your data resides** and **connecting it together**. Microsoft **OneLake** is an *AI-ready data lake* built to unify your entire multi-cloud data estate. Similar to how OneDrive provides a single, central place for files in Microsoft 365, **OneLake** serves as a single, centralized hub for **all** enterprise data.

Most data estates don't live in one place. They're spread across multiple clouds, accounts, databases, domains, and data platforms. Data engineers end up spending enormous effort stitching these systems together with pipelines, moving or copying data, and trying to keep everything in sync. OneLake simplifies this through a **zero-copy, zero-ETL approach.**

Whether your data is in Azure, Amazon Web Services, Google Cloud, on-premises SQL Server, or platforms like SAP, Dataverse, Snowflake, or Databricks – you can connect it to OneLake in just a few clicks, **without moving or duplicating it**.

No more sprawling ETL pipelines. No more out-of-date copies. No more data silos. Once your data is connected to OneLake, you only need a single copy across every Fabric engine or even other engines in Snowflake or Azure Databricks.

> **Customer Highlight:** Using OneLake shortcuts, mirroring, and Direct Lake mode, **Lumen**, a global enterprise technology company, eliminated over 10,000 hours of manual data copying. "We used to spend up to six hours a day copying data into SQL servers," says Chad Hollingsworth, Cloud Architect at Lumen. "Now it's all streamlined… OneLake allowed us to ingest once and use anywhere

Once your data is connected to OneLake, it becomes easily discoverable through the OneLake catalog, where data assets are listed along with their governance and security metadata. The OneLake catalog is integrated into the tools your teams already use—such as Power BI, Microsoft Teams, Microsoft Excel, Microsoft Copilot Studio, and Microsoft Foundry—so users can find and use data where they work. OneLake is even natively integrated with Foundry IQ, making it simple to connect custom agents directly to your multi-cloud data sources.

Processing and harmonizing data with Fabric Analytics

AI agents are **only as reliable as the data you feed them**. Before enterprise data can train or ground an AI agent, the data often needs to be **cleaned, curated, and validated** so the agent is working from consistent, trusted information. That foundation must also extend to real-time data. Agents need to continuously sense what is happening across systems and environments, understand events in context, and respond fast enough to influence outcomes. Without governed, high-quality historical and real-time data, agents simply inherit fragmentation, lag, and quality gaps, leading to unreliable or outdated decisions.

Fabric provides industry-leading analytics engines to help your data teams prepare data for AI. These include tools for data ingestion and orchestration with **Fabric Data Factory**, big data processing with **Spark in Fabric Data Engineering** and **T-SQL in Fabric Data Warehousing**, machine learning and modelling in **Fabric Data Science**, and business intelligence with **Power BI**. For live signal data, **Fabric Real-Time Intelligence** enables continuous streaming, analysis, modelling, visualization, and action. This allows agents and users to reason over current conditions and make operational decisions as events unfold.

Crucially, all these experiences are **unified in Fabric** and operate over OneLake. This means as data moves through pipelines, notebooks, warehouses, or streaming jobs, it stays in OneLake as a single copy—governed, discoverable, and continuously available to power downstream analytics, reporting, and AI applications.

Creating semantic meaning with Fabric IQ

The next leap comes from adding **context** to your data. That's where **Fabric IQ** comes in.

The power of Fabric IQ lies in how it unifies disparate data types like telemetry, time series, graph, and geospatial under **a single, coherent semantic framework called an ontology**. Generated from your existing Power BI semantic models, ontologies organize operational and analytical data together through business entities, relationships, properties, rules, and actions.

Instead of thinking in terms of tables and schemas, your teams and agents can operate on this framework, or ontology, aligned to **how the business actually operates**.

With this semantic foundation in place, organizations can build operations agents in Fabric that function like virtual team members: Continuously monitoring real-time data, detecting patterns or anomalies, and taking proactive action based on predefined business logic. You can also use **planning in Fabric IQ** to create plans, budgets, forecasts, and scenario models directly on top, giving you a complete, contextual view of your historical, real-time, and forward planning data. This allows users and agents to quickly answer what has happened, what is happening, and what should happen all from a single source.

In short, Fabric IQ is designed to **model reality with data**, so that every insight, prediction, and action taken by an agent is grounded in the context of your business.

Empowering AI agents to act with Fabric data and operations agents

Leading-edge organizations are moving beyond the idea of a single, general-purpose AI assistant. Instead, they are adopting **multi-agent systems** composed of specialized agents, each grounded on specific data and performing specific roles. Working together, these teams of agents deliver outcomes that are more accurate, trustworthy, and scalable than a one-size-fits-all approach.

The foundation of any scalable multi-agent strategy is the ability to create **reusable agents** that are connected to trusted enterprise data. Microsoft Fabric makes this possible by allowing you to build agents *within* the data platform itself.

Fabric data agents act like virtual data analysts, aligned to business domains and able to reason across all data in OneLake, while **Fabric operations agents** monitor the data estate and take action in real time, as described above. These agents can be used natively within Fabric experiences, or they can serve as foundational knowledge or capability providers to other AI platforms. For example, you could plug a Fabric data agent into Microsoft Foundry, Copilot Studio, or even Microsoft 365 Copilot, to provide richer answers or analysis.

> **Customer Insight:** *"Microsoft Fabric data agents form the conversational layer. Using Fabric and Microsoft Foundry, we've built agents that allow us to talk to our data, ask personalized questions, and uncover insights based on role or function."*
>
> *— Maureen Tan, Head of AI Center of Excellence, NTT DATA.*

Securing and governing your data and agents

As employees and AI agents increasingly rely on data to make decisions, secure data sharing becomes essential. Overly restrictive access often leads to shadow datasets and workarounds that reduce transparency and increase risk. Microsoft Fabric and OneLake enable broad data access while protecting sensitive information through built-in encryption, strict access controls, and fine-grained permissions that are consistently enforced across agentic and non-agentic experiences. Fabric also integrates with Microsoft Purview data security including *Data Security Posture Management* which provides a centralized control plane to monitor interactions with Fabric data agents and Copilot in Fabric. This means your agent activities can be governed with the same level of rigor, transparency, and trust as human access.

Implementation checklist: Fabric IQ

Plan

- **Inventory your data estate** across clouds, platforms, and business domains to identify which data sources will be needed to train or ground your agents. Learn more
- **Create shared definitions of business entities** like suppliers, products, and routes for your shipping department, so agents can reason over them and identify where they can be most effective at automating tasks and decisions. Learn more
- **Establish data governance and security principles early on** – Decide what data access controls, sensitivity labels, and lineage tracking you will require for AI scenarios. Learn more

Build

- **Unify your data estate** using shortcuts and mirroring in OneLake to enable zero copy, multi cloud access without duplicating data. Learn more
- **Process and refine historical and real-time data** using Fabric's analytics engines and Fabric Real-Time Intelligence so that data is clean, consistent, and AI-ready. Learn more

- **Create ontologies** on top of your Power BI semantic models to define business entities, relationships, properties, and actions as the foundation of Fabric IQ. Learn more

Deploy

- **Manage ontologies in Fabric IQ for your key business areas** so that agents interpret data using business language, rules, and relationships – Not raw technical tables. Learn more
- **Build Fabric data agents and operations agents** aligned to specific business domains to provide conversational analysis and take autonomous and semi-autonomous action. Learn more
- **Integrate agents into downstream experiences** such as Microsoft Foundry, Copilot Studio, or Microsoft 365 Copilot to power multi agent solutions. Learn more

Operate

- **Use the OneLake catalog** to manage discovery, access, and oversight of data across the organization. Bring data into the apps people use every day like Teams, Excel, and Power BI, to make discovery and governance part of the normal workflow. Learn more
- **Enforce fine-grained OneLake security** and use Microsoft Purview sensitivity labels to ensure that as agents and users consume data, all access is compliant with your policies. Learn more

Foundry IQ: Unlocking knowledge for your agents

As AI agent adoption grows, organizations must deliver applications faster and with higher quality while keeping their data secure. This is where **Foundry IQ** comes in. Foundry IQ provides a **context engineering platform** for Microsoft's intelligence layer, unlocking knowledge for your agents, wherever it lives. It **automatically handles how agents connect to, process, and action enterprise information**. With built-in enterprise security and agentic retrieval, agents deliver better results from the right context. Foundry IQ enables agents to activate knowledge from everywhere for faster agent delivery, use context without blind spots to unlock better results, and run on enterprise-grade security, with identity and policy built-in.

Foundry IQ is anchored on three core pillars:

- **Activate knowledge from everywhere:** Faster agent delivery with point and click knowledge bases, reusable across agents.
- **Context without blind spots:** Unlock better results with agentic retrieval that finds the relevant data automatically
- **Enterprise-grade security** with identity and policy built-in. Every answer respects your organization's security, identity, and compliance policies by default

Activate knowledge from everywhere

Foundry IQ enables faster app delivery with point and click knowledge bases, reusable across agents.

It **fuels agents with enterprise-wide knowledge from everywhere** by grounding agents with unified context across the enterprise: Work IQ, Fabric IQ and virtually everything with Foundry IQ.

Knowledge can be reused across agents and apps. Build a knowledge base for a topic once, then plug it into multiple agents and applications to eliminate duplicate data wiring and maintain development velocity.

Break down data silos with point-and-click data transformation and embedding. Minimize upfront data preparation with automated, incremental indexing pipelines tuned per data type, offloading data preprocessing from the application with industry-leading techniques across the data pipeline.

Build with data lakes and custom apps in mind. Automatically store metadata associated with structured and semi-structured knowledge into an index usable for faceted search and other AI app use cases with no additional effort.

Context without blind spots

Unlock better results with agentic retrieval that finds the relevant data automatically.

Use agentic AI to deliver superior context. Foundry IQ's agentic RAG engine plans, source selects, searches, responds and iterates to ensure your agent gets the best answer without wasting tokens. Powers M365 Copilot and available in frameworks like Llamaindex, Langchain, Semantic Kernel, AutoGen, Haystack, n8n and mem0.

Build with best-in-class RAG performance at every phase of AI maturity. Improve response performance by delivering 36% better quality than traditional RAG engines.

Automate retrieval performance for latency or quality. Activate the retrieval engine that meets your specific goals between latency and quality with retrieval reasoning effort. Let the system adapt on its own or fully customize your RAG stack for more control.

Full context with expansive reach of data. Whether it's multimodal, indexed, remote, Foundry IQ understands your data and how it's related, delivering integrated context that is hard to achieve with siloed tool calling.

Enterprise grade security, with identity and policy built-in

Every answer respects your organization's security, identity, and compliance policies by default.

Integrated user permissions for secure AI. Enforce document-level access control at query time by honoring ACLs indexed from Microsoft 365 and other sources, with Entra ID based identity propagation ensuring AI applications and agents retrieve only content each user is authorized to access.

Enforce unified governance automatically, ensuring safe and compliant answers. Purview integrated sensitivity labels are extracted during indexing, enforced at query time, and preserved end-to-end across SharePoint, OneLake, Blob, and ADLS Gen2. SharePoint permissions, shared links, ACLs, and labels flow through unchanged for remote SharePoint content—without extra setup or policy duplication.

Enterprise-grade network isolation for AI workloads. Foundry IQ runs in your own virtual network with Private Endpoints, Azure Private Link, and managed identities, supporting agentic and knowledge retrieval with zero public internet exposure

Implementation checklist: Foundry IQ

Plan

- Identify the app scenarios where agents will need **curated knowledge from multiple sources and types**, not just live work context or metrics.
- Define what content belongs to a particular domain or persona for each app scenario, and identify where that content lives (e.g. HR policies, IT helpdesk admin, engineering wiki pages, customer support agent).
- Inventory your knowledge sources: List out relevant SharePoint sites, OneDrive folders, OneLake locations, databases, wiki pages, intranet sites, documentation repositories, webpages, reports etc.
- Document any security and governance requirements specific to these knowledge sources

Build

- Prerequisites: Azure AI Search service, Microsoft Foundry project with an LLM deployment, authentication and permissions on your search service and project.

- Authentication and permissions: Role-based access control is recommended for production deployments. If roles aren't feasible, use key-based authentication.
- Create a knowledge base aligned to a domain topic.
- Add relevant knowledge sources to your knowledge base.
- Configure the agentic retrieval settings for your scenario if needed: For example, select the model that will be used in the query engine and set retrieval reasoning effort based on performance and latency priorities.
- Connect a knowledge base to a Microsoft Foundry project. https://learn.microsoft.com/en-us/azure/foundry/agents/how-to/foundry-iq-connect?tabs=foundry%2Cpython
- Create and connect an agent with MCP support to the Foundry project and knowledge base.

Deploy

- Validate **the retrieval reasoning effort and permission-aware behavior** by testing queries via API or MCP endpoint. Confirm an agent only returns content that each user is allowed to see, and learn how responses are delivered on the different retrieval reasoning settings. https://learn.microsoft.com/en-us/azure/search/agentic-retrieval-how-to-retrieve
- Enable **tracing and observability** on your agent (via Foundry or your framework) so you can monitor what sources the agent consulted and how it formulated its answer.
- Test complex, multi-hop questions and **review the activity array** to ensure the agentic retrieval is operating as expected.

Operate

- Assign clear ownership for knowledge base costs and operations.
- Enforce enterprise governance on the knowledge bases: For example, require approvals before adding a new source or ensure compliance review for content added.
- **Monitor and improve retrieval quality** – Incorporate additional indexing and relevance-tuning features to improve the quality of generated results, like scoring profiles and semantic configuration.
- Reuse knowledge bases across multiple agents and applications; take advantage of that by connecting new agents to existing knowledge bases whenever appropriate.

Kickstart the journey with Microsoft Agent Factory

Scaling from experimentation to full production requires more than tools—it requires a **disciplined operating model** and shared patterns for how agents are built, deployed, and governed. Microsoft Agent Factory is designed to help organizations bridge this gap.

Below is a **recommended first 90-day adoption path** that builds confidence while establishing the foundations needed for scale:

01 Activate a unified licensing and skilling plan

As part of Agent Factory, organizations can adopt a single, pre-paid plan (SKU) that covers building agents across Microsoft Copilot Studio, Microsoft Foundry, GitHub Copilot, and Microsoft Fabric. This gives teams flexibility to choose the right tool for each use case under one budget with more predictable costs. The plan also includes **tailored, role-based skilling** to accelerate adoption and build confidence, at no additional cost. Explore the Microsoft Agent Prepurchase Plan.

02 Start with Work IQ scenarios

Begin with high-value, low-friction agent scenarios that operate **in the flow of work** within Microsoft 365. For example, agents that assist with drafting content, triaging incoming requests, or preparing for meetings. These leverage Work IQ context and can deliver immediate productivity gains while staying within existing identity, permission, and compliance boundaries. These early wins help teams learn in a safe context. Learn more by exploring the following resources: Documentation, Microsoft 365 Copilot >Work IQ, Microsoft 365 Enterprise > AI.

03 Establish a unified data foundation with OneLake

Establish a unified data foundation by connecting your distributed data estate into Microsoft OneLake as a centralized catalog. Use shortcuts and mirroring to enable a zero-copy, zero-ETL architecture, allowing data across clouds and platforms to be accessed and synchronized without duplication or complex pipelines. By centralizing

access in this way, you eliminate silos and stale copies while creating a single, consistent data layer that all analytics and AI experiences can rely on. Explore the documentation.

04

Prepare your data, add context, and create data agents

Prepare and harmonize your data using **Microsoft Fabric** analytics engines so it is clean, structured, and AI-ready. Establish a semantic foundation with Fabric IQ to organize data around business entities, relationships, and rules, enabling consistent, context-rich understanding across analytics and AI. Build **domain-specific data and operations agents** on top of this foundation, allowing them to reason over data and take real-time action, while integrating with platforms like Microsoft Foundry, Microsoft Copilot Studio, and Microsoft 365 Copilot to support multi-agent systems. Explore the documentation.

05

Create a reusable Foundry IQ knowledge foundation

Build a shared **Foundry IQ knowledge base** that spans some of your most important text and document sources (e.g. policy documents, support knowledge, technical manuals). This will serve as a unified, **permission-aware grounding layer with citations** that both low-code and pro-code agents can draw from. By doing this centrally, you avoid having each team or project create its own document index. It reduces duplicate effort and improves answer quality across all agents, since they'll be drawing from a common, consistently managed body of knowledge. Explore the documentation.

To further accelerate your AI onboarding, engage Microsoft partners to implement Microsoft IQ.

Why this Sequence Works

This staged progression allows organizations to:

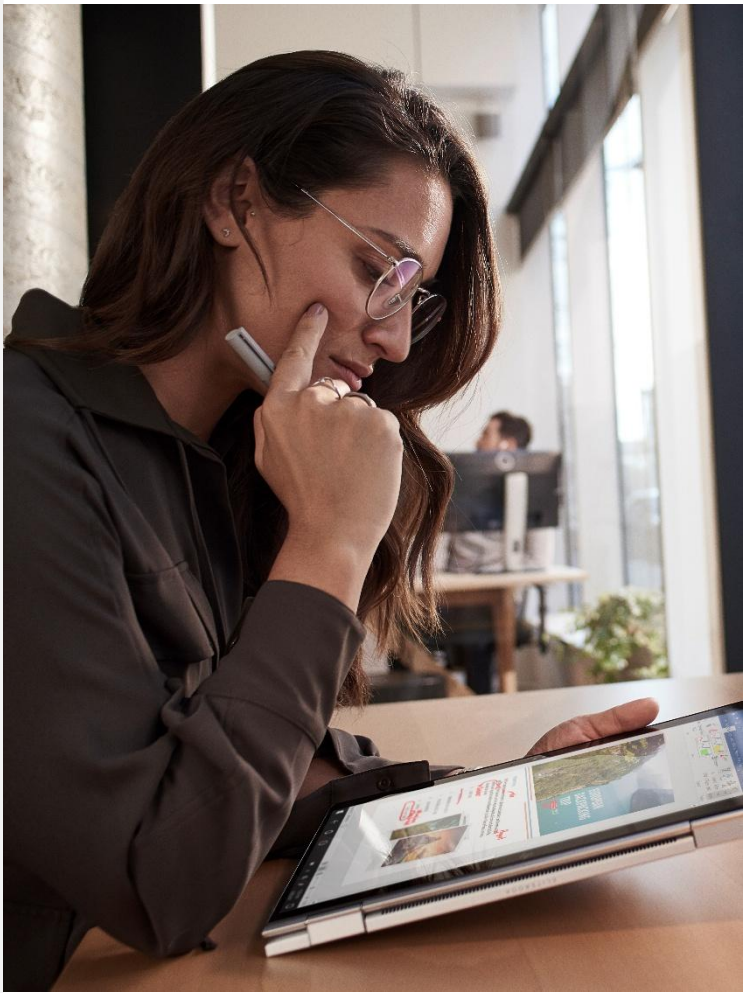
- Deliver early business value **quickly** without bypassing any security or compliance standards (initial agents stay within well-understood Microsoft 365 contexts).
- Establish shared data and knowledge foundations **before** significantly increasing agent autonomy or scope (so later agents all build on a solid, consistent base).
- Introduce real-time capabilities **gradually** in areas where they provide clear benefit, rather than overwhelming the system with real-time demands everywhere.

By following this approach, teams avoid creating a collection of disjointed "one-off" agents or siloed pilot projects. Instead, you build toward a **cohesive enterprise agent platform** where each new agent incrementally adds to a well-governed, scalable ecosystem.

If your organization needs additional support in this journey — for instance, to build production-ready, full-stack AI solutions — Microsoft Agent Factory can help you tap into deep expertise. This might include Microsoft's **Forward Deployed Engineering (FDE)** resources or certified partners who can work alongside your teams. The end goal is to move confidently from early **pilots to an enterprise-wide platform** for agents, without sacrificing trust, control, or operational excellence.

From Pilots to Platform

Agent Factory isn't about accelerating one or two agents. It's about helping your organization build the **capability** to produce, deploy, and manage AI agents repeatedly, safely, and at scale. By combining the Microsoft IQ architecture, governed runtimes, and integrated security/operations, your organization can move from initial experiments to a production-grade, enterprise-wide agent platform. In doing so, you can embrace AI agents as a lasting, strategic part of your business **without compromising on trust, control, or operational rigor**.



Consolidated Implementation Checklists

Work IQ

- **Build path:** Options for creating Work IQ-driven agents include using built-in Copilot experiences, Copilot Studio (for custom internal agents), or integrating Work IQ context into your own agent frameworks. Work IQ blog
- **Connecting external content:** Use Microsoft 365 Copilot *connectors* (for federating external data into Copilot) and Power Platform connectors to include third-party or on-premises data in your Work IQ scope. Copilot connectors docs
- Begin to build **business understanding** – For instance, define key business entities and actions via Dataverse so agents can reliably interact with them. Learn More

Fabric IQ

- **Plan your data and governance foundation in OneLake:** Inventory data, define key domains, align stakeholders, and set a governance and security. OneLake overview.
- **Unify and prepare your data:** Connect to your data across your estate in OneLake, then clean and model it into AI-ready semantic structures. Analytics tutorials in Fabric.
- **Build and deploy domain-specific data agents:** Create ontologies and agents for key domains, and integrate them into platforms like Foundry, Copilot Studio, and Microsoft 365 Copilot. Fabric IQ overview.

Foundry IQ

- **Point and click knowledge bases, reusable across agents:** Create Foundry IQ knowledge bases to handle all components for a domain topic, including relevant knowledge sources, agentic retrieval orchestration and permissions at document and user level. What is Foundry IQ?
- **Connect Foundry IQ to an agent in Foundry:** Connect knowledge bases to your agents directly from the Foundry portal. Connect Foundry IQ knowledge base
- **Agent tracing for debugging:** Enable agent tracing and logging to debug agent behavior and monitor the knowledge base's activity array to ensure retrieval steps are performing as expected. Tracking and monitoring agentic retrieval outputs