



Work IQ before bed

*A slow walk through Work IQ,
the part of Microsoft IQ that knows how you work*

BY DIEGO · OCTOBER 2026 · 20 MINUTES

BEFORE YOU START

How to read this

Written for the end of a long day

One idea per chapter. Short lines. No code.

Every chapter ends with one line to remember. If you're too tired for the chapter, read only that line.

Technical words get explained the moment they show up.

Facts are checked against Microsoft Learn as of October 2026.

Several parts of Work IQ are still in preview. The book says which ones.

It's still ok.

CHAPTER 1

The Tuesday problem

Why this thing exists

It's Tuesday. A customer call starts in ten minutes.

The context is everywhere. Two email threads. A Teams chat where someone promised a date. A deck in SharePoint, version seven. Last week's meeting notes. A colleague who knows the history but is on holiday.

You know where to look. Mostly. An AI agent doesn't. Give it a great model and no access to your work and it's a very well-read stranger.

Work IQ is the layer that lets Copilot and agents understand your work, inside your permissions.

Microsoft calls it a workplace intelligence layer. In plain words: it reads the signals of how work happens in Microsoft 365 and builds an understanding of them. Then it gives agents a safe way to ask and act.

It became generally available on 16 June 2026. Several pieces around it are still in preview. We'll flag them as we go.

REMEMBER

Work IQ gives agents your work context. Only what you're allowed to see.

One of four IQs

Microsoft IQ and where Work IQ sits in it

Work IQ is one part of **Microsoft IQ**, the enterprise intelligence layer Microsoft puts under Copilot and agents. Four parts, each with its own kind of knowledge.

Work IQ how people work: mail, meetings, chats, files

Fabric IQ how the business runs: data and its meaning

Foundry IQ what the company knows: policies, documents

Web IQ what the world says: fresh web results

Four kinds of context, one intelligence layer.

If you read the Fabric and Foundry books, two of these are old friends. Fabric IQ describes business data in business terms, with ontologies. Foundry IQ builds knowledge bases from documents for agents.

Work IQ is the human side. Who talked to whom. What was decided in which meeting. Which file is the latest one.

Next to them sits **Agent 365**, the registry and control plane for agents. It keeps track of which agents exist and what they're allowed to do.

Each IQ lives in its own product. Work IQ comes with Microsoft 365. Fabric IQ lives in Microsoft Fabric. Foundry IQ lives in Microsoft Foundry. Web IQ brings fresh results from across the web, so an agent isn't stuck with what a model learned during training. An agent can use one of them or several. The good ones use whichever fits the question.

REMEMBER

Work IQ is the people-and-collaboration part of Microsoft IQ.

Why not build it yourself?

The question every developer asks first

Every developer's first reaction is the same. I can call the Microsoft 365 APIs myself. Why do I need a layer?

You can. Here's what that project looks like.

You copy mail, chats and files into a search index of your own. You turn the text into vectors, lists of numbers that capture meaning, so you can search by meaning. You write sync jobs to keep the copy fresh. Then you rebuild Microsoft 365's permissions on top of your copy, so nobody sees a file they shouldn't. Then you add compliance rules. Then you discover a deleted email still answering questions in your index.

That's months of work. Most of it is plumbing and risk.

Work IQ skips the copy: answers come from the data where it lives, with its permissions already on.

Microsoft lists three reasons in the docs. Secure grounding, because access is permission-trimmed and policy-enforced automatically. Flexible ways to interact, from finished answers to raw tools. And less to operate: no vector stores, no data sync jobs, no home-made compliance.

Building your own still makes sense for unusual data or very special search needs. For "what's happening in my work", it rarely does anymore.

REMEMBER

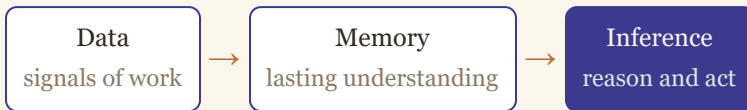
No private copy, no sync jobs, no rebuilt permissions.

CHAPTER 4

Data, memory, inference

The three layers inside

Microsoft describes Work IQ as three layers that work together.



From raw signals to an agent that can act.

Data brings together the signals: files, emails, meetings, chats and business applications. It's the raw record of how work happens.

Memory builds a lasting understanding of how people and teams work, so agents stay aligned with priorities across tasks, apps and sessions. The agent doesn't start from zero every morning.

Inference brings models, skills and tools together, so agents can reason and then do something. Read, draft, schedule, update.

An everyday comparison. Data is the filing cabinet. Memory is the colleague who has been around for years and knows which drawer matters. Inference is that colleague actually getting up and doing the task.

REMEMBER

Data is the record, memory is the understanding, inference is the action.

What it can see

The sources behind the answers

Work IQ reasons over the Microsoft 365 data you already have.

Email. Messages in Outlook.

Meetings and calendar. What's planned, who's invited, what happened.

Documents. Files in OneDrive and SharePoint.

Teams. Chats and channel messages.

People. Who works with whom. The org chart.

Planner. Plans and tasks.

Enterprise search. Results from the organization's search.

Then there's the business side. **Business Applications in Work IQ**, in preview, adds data from Dynamics 365 and Power Platform through Dataverse, the database under those apps. Opportunities, cases, orders. So "how many deals close next month?" can be answered next to "what did the customer say in last week's email?".

That part comes with **business skills**, reusable descriptions of business processes. It also uses **semantic models** that tell Work IQ what the business tables mean.

The questions Microsoft uses as examples are refreshingly ordinary. What did John say about the proposal? What's on my calendar tomorrow? Who is working on Project Alpha? How many opportunities are closing next month? Nobody needs an AI to be brilliant here. They need it to know where things are.

REMEMBER

Mail, meetings, files, chats, people, plans and search. Business apps in preview.

CHAPTER 6

Chat, context, tools, workspaces

Four things Work IQ offers to builders

For people building agents, Work IQ comes as four building blocks.

Chat gives you finished answers. You ask a question and get a fully processed, Copilot-quality response. Good when you want the result, not the raw material.

Context gives you the raw material, ready to use. Work IQ gathers and grounds the relevant pieces itself, so your agent doesn't have to stitch signals together or run its own search pipeline.

Tools let agents act. A small set of generic tools reads and changes Microsoft 365 data: fetch an email, create a meeting, update a file. Chapter eight explains the trick behind them.

Workspaces give long-running agents a place to keep their work in progress. It's SharePoint Embedded storage inside your own tenant, so results can be reused and handed over to the next agent.

Chat when you want the answer, context when you want the facts, tools when you want something done.

REMEMBER

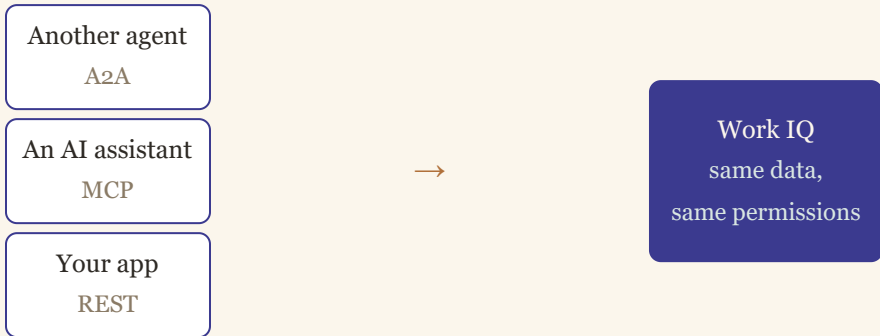
Chat answers, context informs, tools act, workspaces remember the job.

CHAPTER 7

Three doors in

A2A, MCP and REST

Different callers need different doors. Work IQ has three.



Pick the door that matches who is calling.

A2A (Agent2Agent) is for agents talking to agents. Your agent hands Work IQ a task, like "find what we promised Contoso last month". It gets the result back. Work IQ acts as a peer agent. This is how a Foundry agent connects to it.

MCP (Model Context Protocol) is the standard plug that lets an AI assistant use tools. Work IQ offers a remote MCP server in the cloud. There's also a local one that runs on your machine. Coding assistants like GitHub Copilot use this door.

REST is the classic web API. Your app sends a question and shows the reply. No agent framework needed.

Conversations can continue. With A2A, each answer comes back with a conversation identifier. Send it with the next question and Work IQ keeps the thread. "Tell me more about the two o'clock call" then means the call from the previous answer.

One small detail that bites: time. "What meetings do I have today?" depends on where you are. A2A requests carry the user's time zone. Microsoft's docs say it's required for time-sensitive questions.

REMEMBER

A2A for agents, MCP for assistants, REST for apps.

Fewer tools, more paths

The clever part of the MCP design

Microsoft 365 has thousands of operations. Give an agent one tool per operation and it drowns. Too many choices, too many tokens, too many mistakes.

Work IQ does the opposite. A small set of generic tools act as verbs. The **resource path** says what they act on. A path is an address, like a folder path on your laptop.

fetch /me/messages. Read my emails.

create_entity /me/events. Create a meeting.

do_action /me/sendMail. Send an email.

ask. Hand a question to Microsoft 365 Copilot.

New data sources add new paths. The tool list stays the same size.

And agents don't need to know every path in advance. A schema tool lets them ask at runtime what exists and how it's shaped.

Microsoft calls this introspection over enumeration. Ask instead of memorizing.

How many tools exactly? The overview page says ten. The tool reference lists twelve today, including a **retrieve** tool still in preview. Not a big deal. The number is small and that's the point.

The tools are verbs. The paths are nouns. The list of verbs never grows.

REMEMBER

A few generic tools plus paths. Schemas discovered at runtime.

Only what you may see

Permissions, policy and logs

Every Work IQ request runs as a specific user. It only reaches what that user is allowed to see or do. If you can't open a file, an agent asking on your behalf can't either.

The old way to secure APIs was hundreds of narrow permissions, called OAuth scopes. One for reading mail, one for calendars, one for each kind of file. Hard to manage and easy to get wrong.

Work IQ uses a handful of broad permissions to set the outer boundary. Then a **policy engine** checks every single request. It uses Rego, the policy language of Open Policy Agent, an open-source standard. It looks at the path, the method, the user and even the data itself. Then it decides yes or no.

Every tool call is logged and evaluated. That gives audit trails, usage numbers, rate limits and compliance checks in one place. Admins manage it in the Microsoft 365 admin center.

When agents act, the Agent 365 control plane keeps those actions observable and governed. For business data, admins can review which users, agents and applications have access. So the question "which agent touched our customer records last week?" has an answer. Before, it mostly had a shrug.

There's a flip side. Permission-trimmed means an agent finds everything you could find yourself. If a sensitive file is shared too widely in SharePoint, Work IQ can surface it. Fix the sharing, not the agent.

REMEMBER

User-scoped, policy-checked, logged. Oversharing in SharePoint is still your problem.

Where you'll meet it

The front doors people actually use

Microsoft 365 Copilot. Work IQ is already inside Copilot experiences and agents for licensed users. Many people use it without knowing the name.

Copilot Studio. Add Work IQ (preview) to an agent as one tool surface for Microsoft 365. It runs on the GitHub Copilot harness and uses Copilot Credits.

Microsoft Foundry. A Foundry agent connects to Work IQ (preview) over A2A and delegates questions about the user's mail, meetings and files.

Work IQ CLI. A command-line tool. Type *workiq ask* with a question in a terminal, or run it as a local MCP server so a coding assistant can use your work context.

Cowork. Microsoft 365 Copilot's Cowork has a Work IQ plugin.

One trap in Copilot Studio. The tool catalog also shows older, separate tools for Mail, Teams, Calendar, SharePoint and others. Those are the legacy experience, kept for compatibility. They aren't Work IQ under an old name. Different runtime, different billing. Pick Work IQ (preview) for the new experience.

I use the CLI myself. The desktop assistant I work with calls it to answer questions about my week. Most days it's faster than me remembering which chat something was in. Not a high bar.

REMEMBER

Copilot, Copilot Studio, Foundry, the CLI and Cowork. Same Work IQ underneath.

The CLI on your desk

The fastest way to feel what Work IQ does

The Work IQ CLI is the shortest path from curiosity to an answer. A CLI is a command-line interface: you type a command in a terminal window and get text back.

It works in two modes.

CLI mode. You type *workiq ask* followed by a question in plain language. "What's on my calendar tomorrow?" "Find my recent documents about Q4 planning." "Summarize today's messages in the Engineering channel and propose a plan of action." The answer comes back in the terminal.

MCP server mode. The CLI runs quietly as a local MCP server. Your AI coding assistant, like GitHub Copilot in Visual Studio Code, can then pull in work context while you code. You're building the feature discussed in yesterday's meeting. The assistant can read that meeting's context and suggest code that matches what was agreed.

It runs on Windows, Linux and macOS. It installs with npm, the package manager that comes with Node.js. You need a usage-based billing plan and an admin's consent first. Chapter thirteen covers that.

If you only try one thing after reading this book, try this. Ask it something you'd normally spend ten minutes searching for. Then decide what you think.

REMEMBER

workiq ask in a terminal, or a local MCP server for your coding assistant.

Who pays

Licensing in one page

Work IQ API access doesn't depend on the Microsoft 365 Copilot license. It's billed by usage.

Copilot users. Work IQ is available in all Microsoft 365 Copilot experiences and agents. Custom and third-party agents that call it are billed by usage.

Everyone else. Billed by usage.

Usage is paid in **Copilot Credits**, the pay-as-you-go currency also used by Copilot Studio. To use it, an organization sets up a usage-based billing plan in Copilot Studio, linked to an Azure subscription and resource group. Then it assigns users to it.

Admins watch and cap the spend in the Microsoft 365 admin center.

The license covers Copilot. The meter covers your own agents.

That split matters when a team builds its first custom agent. The pilot is cheap. A popular agent calling Work IQ all day is a real line on the bill. Set the budget before the launch, not after.

REMEMBER

Usage-based, in Copilot Credits, managed in the admin center.

Turning it on

The five-minute admin part

Setup is short. It does need the right people.

- 1. A billing plan.** Usage-based billing in Copilot Studio, linked to an Azure subscription and resource group.
- 2. Users on the plan.** The people who will use Work IQ are assigned to it.
- 3. A Global Administrator, once.** They create the Work IQ service principal in the tenant. A service principal is the identity an app uses in Microsoft Entra. Microsoft estimates about five minutes.
- 4. Admin consent for the CLI.** The CLI needs permissions only an admin can grant for the organization.

Normally the service principal appears by itself the first time anyone uses Work IQ. One known corner case: an admin who sets Work IQ policies in the admin center before anyone has used it can hit an error, because the identity doesn't exist yet. The fix is step three.

REMEMBER

Billing plan, users, one admin step, consent for the CLI.

Tuesday, again

The three books, in one story

Back to the customer call in ten minutes. This time an agent in Foundry prepares the briefing.

1. The user asks: "Brief me for the Contoso call."
2. The Foundry agent hands Work IQ a task over A2A: last month's emails, chats and meetings with Contoso.
3. Work IQ runs as the user. It finds the promised delivery date in a Teams chat and the open question in an email. Things the user can't see stay hidden.
4. The agent asks a Fabric data agent for consumption numbers. Fabric IQ knows what "consumption" means for this customer.
5. Foundry IQ supplies the contract policy from the legal knowledge base.
6. The agent writes a one-page briefing and saves it in a Work IQ workspace, for the follow-up agent after the call.
7. Every Work IQ call is logged and policy-checked. Copilot Credits tick by. Agent 365 knows the agent exists.
8. After the call, a follow-up agent opens the same workspace. It reads the briefing, adds what was agreed and drafts the recap email. Nobody asks the same questions twice.

9. Next week the user asks about Contoso again. Memory remembers this is a priority account. The answer starts where last week's ended.

Work IQ for the people. Fabric IQ for the numbers. Foundry IQ for the rules.

The briefing takes a minute. The call goes better. Probably.

REMEMBER

Work IQ brings the human context the other IQs don't have.

Things I watched people trip on

Including me

Picking the legacy tools. In Copilot Studio, the old Mail or Teams tools aren't Work IQ. Different runtime, different billing.

Preview in production. Work IQ in Copilot Studio, in Foundry and for business apps is in preview, as is the retrieve tool. Fine for pilots. Check before a production contract.

Forgetting the meter. Custom agents are billed by usage, even for Copilot-licensed users. Budget first.

Oversharing. Agents see what users can see. A messy SharePoint becomes a messy answer, or worse.

Compliance boundaries. Microsoft's Foundry docs warn that data can leave the Azure compliance boundary when an agent calls Work IQ. Decide where data may flow before you connect.

Time zones. "Today" depends on where the user is. Send the time zone with time-sensitive questions.

Waiting for an admin. Billing plan, service principal and consent all need admin rights. Line them up before the workshop.

Expecting magic on a messy tenant. Answers get better with clear file names, named owners and current versions.

Seven copies of "final_v3" confuse people and agents alike.

Forwarding without checking. Ask where an answer came from and open the source before a number goes to a customer. The agent is fast. Your name is on the email.

The pocket glossary

For when you wake up and forgot half of it

Work IQ. The workplace intelligence layer behind Copilot and agents.

Microsoft IQ. Work IQ, Fabric IQ, Foundry IQ and Web IQ together.

Data, memory, inference. Work IQ's three layers: record, understanding, action.

Chat. Finished answers from Work IQ.

Context. Grounded raw material for your own agent.

Tools. Generic verbs that read and change Microsoft 365 data.

Workspace. SharePoint Embedded storage for an agent's work in progress.

A2A. Agent2Agent, the protocol for agents handing tasks to agents.

MCP. Model Context Protocol, the standard plug for AI tools.

REST. A classic web API for apps.

Resource path. The address of the data a tool acts on.

Introspection. Asking for a schema at runtime instead of memorizing it.

OAuth scope. A narrow permission an app requests.

Rego. The policy language Work IQ uses to check each request.

Business Applications in Work IQ. Dynamics 365 and Power Platform data, in preview.

Copilot Credits. The usage-based currency Work IQ is billed in.

Service principal. An app's identity in Microsoft Entra.

Work IQ CLI. Command-line access. Also a local MCP server.

Agent 365. The registry and control plane for agents.

LIGHTS OUT

Before you close it

Fabric knows the numbers.

Work IQ knows the people behind them.

Go to sleep.

TL;DR

Work IQ gives Copilot and agents the context of your work, inside your permissions.

It's one of four IQs. Work for people, Fabric for business data, Foundry for documents, Web for the web.

Three layers: data, memory, inference.

Four building blocks: chat, context, tools, workspaces.

Three doors: A2A for agents, MCP for assistants, REST for apps.

A few generic tools act on paths. Schemas are discovered at runtime.

Every request runs as the user, passes a policy engine and gets logged.

Usage-based billing in Copilot Credits. Budget before you launch.

Start with the CLI. It's the quickest way to see what Work IQ can do for you.

Next time: Agent 365, or who keeps track of all these agents.

Sources: Microsoft Learn. Work IQ overview, Work IQ API overview, Work IQ MCP overview, Enable your tenant for Work IQ, Work IQ CLI. Work IQ in Copilot Studio, Work IQ in Microsoft Foundry, Business Applications in Work IQ, Microsoft IQ documentation. Pages updated between June and October 2026. Written in personal capacity. Features change often, check before quoting.