RISK AND GOVERNANCE DEEP DIVE

Microsoft Copilot and Bank Data

Privacy, Security, Compliance, and Use-Risk Assessment

Prepared for U.S. bank leadership, information security, privacy, compliance, legal, audit, and technology teams

Version 1.0 | As of August 20, 2026

Scope: Microsoft Copilot for work, Copilot Chat, Copilot Studio agents, connectors, web grounding, and current model-provider options

This report is a technology and risk-management analysis, not legal advice or a regulatory opinion. Validate conclusions against the bank's contract, tenant configuration, charter, regulators, data locations, and approved use cases.

Document Control and Review Posture

Item	Report position
Primary question	How Microsoft Copilot processes, protects, stores, exposes, and enables governance over bank data and private information.
Bank context assumed	A U.S. insured depository institution using Microsoft 365 commercial cloud, with Copilot Studio considered for selected agents.
Product boundary	Microsoft work offerings are distinguished from consumer Copilot. GitHub Copilot, Dynamics 365 Copilot, Security Copilot, and Azure AI workloads are outside scope unless they enter through an agent or connector.
Evidence posture	Microsoft product documentation is used for product behavior. Statutes, regulations, regulator guidance, NIST, and PCI SSC are used for risk and control expectations.
Decision posture	Enterprise Microsoft Copilot can be conditionally approved after identity, permission, data, retention, web-search, provider, agent, and monitoring controls are validated. Consumer Copilot should not receive bank data.
Review cadence	Revalidate at least quarterly and upon licensing, model-provider, Product Terms, DPA, subprocessor, Purview, Copilot Studio, or tenant-configuration changes.

Most important conclusion

Copilot does not create a new entitlement system. It accelerates retrieval and synthesis across permissions the user already has. The bank's largest practical confidentiality exposure is usually preexisting oversharing, followed by use of the wrong Copilot surface, uncontrolled web grounding, unassessed model providers, and agents that act with maker credentials. [1] [3] [17] [22] [33]

Table of Contents

Executive Summary

Microsoft's enterprise Copilot services can be used in a bank without Microsoft using prompts, responses, or Microsoft Graph data to train foundation models. That assurance is important, but it is only one part of the risk decision. Prompts and responses are normally stored as interaction content, can be subject to retention and legal hold, and can be discoverable by authorized compliance personnel. Copilot also obeys the user's effective permissions, which means that excessive SharePoint, OneDrive, mailbox, Teams, or connector access becomes easier to exercise at speed. [1] [2] [3] [8] [9] [10]

The correct approval unit is not the brand name 'Copilot.' It is the complete combination of product surface, sign-in identity, license, grounding source, web-search state, model provider, agent or connector, credential mode, channel, retention policy, and use case. A bank can have enterprise data protection in one session and materially different treatment in a consumer session, an independent-processor preview model, a third-party agent, or an external web-search flow. [1] [5] [23] [24] [25] [29] [31] [42]

Microsoft remains a service provider and data processor for core enterprise Copilot processing covered by the Microsoft DPA and Product Terms. The bank remains accountable for lawful use, data minimization, access control, classification, retention, third-party oversight, incident response, accurate customer communications, and outcomes. Microsoft compliance certifications and contractual commitments support the bank's control environment, but they do not make a particular use case compliant by themselves. [2] [46] [47] [48] [49]

Decision Summary

Capability	Recommended bank posture	Reason
Microsoft Copilot with work account and enterprise data protection	Conditionally approve	Approve by data class and use case after permissions, labels, DLP, web search, provider, retention, and monitoring tests pass.
Copilot Chat with work account	Conditionally approve	Protected by enterprise terms, but commonly web-grounded. Uploaded or pasted bank content and Outlook scope can introduce organizational data.
Consumer or personal-account Copilot	Do not approve for bank data	Consumer privacy and training controls are different. The enterprise DPA and enterprise data protection commitments do not define this surface.
Copilot Studio read-only agent	Controlled pilot	Approve only with Entra authentication, end-user permission trimming, approved knowledge, restricted channels, transcripts, and testing.
Copilot Studio action agent	High-risk approval	Actions, connectors, maker credentials, external channels, and prompt injection can create transaction and privilege risk.
Autonomous, scheduled, or computer-use agent	Exception only	The agent may act without contemporaneous user review and can use author credentials, user-interface automation, screenshots, and logs.

Capability	Recommended bank posture	Reason
Preview or independent-processor model	Default deny	Terms, retention, geography, audit coverage, and certifications can differ from core Microsoft processing.

Ten Findings That Drive the Risk Assessment

1. **No foundation-model training is not the same as no storage.** Microsoft says prompts, responses, and Graph grounding data are not used to train foundation models. Interaction content is nevertheless stored, retained, audited, and discoverable according to Microsoft 365 and Purview controls. [1] [2] [8] [9] [10]
2. **Permissions are the primary confidentiality boundary.** Copilot returns content the user is authorized to access. It can make old oversharing visible faster, including external-tenant shared content and connector content with inaccurate permission mappings. [1] [3] [17] [20] [22]
3. **The work-account boundary must be unmistakable.** Consumer Copilot has different privacy choices and conversation-history behavior. Bank policy should require the approved work identity and block or prohibit bank data in consumer sessions. [2] [29] [30]
4. **Web search is a separate flow.** Copilot generates a short Bing query rather than sending the full prompt. Microsoft describes identifier removal and limits on advertising and model training, but Bing data is governed under a separate service role and not all Microsoft 365 residency commitments apply. [2] [5] [7] [42]
5. **Third-party model selection is now a first-class control.** In 2026, OpenAI-operated and Anthropic-operated models can participate under different subprocessor or independent-processor configurations. Enablement defaults, geography, retention, certifications, and terms differ. [23] [24] [25] [26] [27]
6. **DLP is valuable but incomplete.** Microsoft Purview can block selected labeled content, sensitive prompt text, external web search, and certain external-email grounding. Directly uploaded files are not content-scanned by the Copilot DLP location, and policy changes can take hours to propagate. [12] [13]
7. **Interaction logs are sensitive records.** Prompts often reveal intent, investigations, personnel issues, customer context, and legal strategy. Compliance personnel may be able to search prompt and response content through Purview, so administrative access needs separation of duties and monitoring. [9] [10] [11]
8. **Copilot Studio materially expands the threat model.** Agents can use knowledge, connectors, HTTP calls, external channels, maker or user credentials, triggers, and automated computer interaction. Each configuration can change the data and authority boundary. [31] [32] [33] [34] [40]
9. **Human review is required for material decisions.** Microsoft warns that generated output can be inaccurate. Banks also must provide accurate and specific reasons for adverse credit action regardless of algorithmic complexity. [1] [51] [52]
10. **Bank obligations remain risk-based and lifecycle-based.** Information-security and third-party guidance require risk assessment, due diligence, contractual protection, testing, monitoring, incident response, and board oversight. The cloud shared-responsibility model does not transfer these duties to Microsoft. [46] [47] [48]

Residual Risk Position

Overall assessment: Moderate after controls

For ordinary productivity use involving public, internal, and selected confidential information, enterprise Microsoft Copilot can reach a moderate residual-risk position when the bank validates tenant controls and user behavior. Customer NPI, legal, HR, cybersecurity, fraud, and financial reporting content require narrower approval. SAR content, authentication secrets, cryptographic material, full payment-card data, and unreviewed automated decisions should be prohibited from general-purpose Copilot surfaces unless a separately approved architecture and legal basis exist. [46] [50] [58] [59]

1. Scope, Terminology, and Assumptions

This report uses 'Microsoft Copilot' as an umbrella term only when a point applies across the named enterprise offerings. Microsoft has changed product names over time, and some administrative pages retain earlier labels such as Microsoft 365 Copilot. The report therefore identifies the functional surface instead of relying on branding alone. [1] [5] [6]

1.1 In Scope

- Microsoft Copilot for licensed work users, including grounding through Microsoft Graph and Microsoft 365 applications.
- Microsoft Copilot Chat for work users, including web grounding, file upload, Outlook context, and agents.
- Microsoft Copilot Studio, including knowledge sources, actions, connectors, channels, event triggers, authentication modes, transcripts, and computer use.
- Microsoft Purview, SharePoint, Entra, and Power Platform controls that materially affect confidentiality, integrity, retention, audit, and access.
- Current OpenAI-operated and Anthropic-operated model options documented for Microsoft Copilot as of the report date.

1.2 Out of Scope

GitHub Copilot, Microsoft Security Copilot, Dynamics 365 Copilot, standalone Azure OpenAI or Azure AI Foundry workloads, embedded third-party AI features, and a bank's own custom machine-learning models are not assessed as products. If any of them supplies data, tools, or model capacity to a Copilot Studio agent, the resulting integration must be assessed separately.

1.3 Key Terms

Term	Meaning in this report

Term	Meaning in this report
Work data	Organizational content available through Microsoft Graph or an approved source, such as email, chats, meetings, files, sites, and business data.
Grounding	Retrieving context and placing it into the model request so the response is more relevant to the user and task.
Prompt	User input plus system instructions, conversation context, retrieved content, files, and agent instructions that may form the effective request.
Enterprise data protection	Microsoft's contractual and technical protections for eligible work-account use under the DPA and Product Terms.
Subprocessor	A provider Microsoft engages to process customer data on Microsoft's behalf under Microsoft contractual obligations.
Independent processor	A provider that processes under its own terms and data agreement, outside the core Microsoft DPA boundary for the feature.
Permission trimming	Returning only source content the current user is authorized to access, assuming source permissions are accurate and supported.
Agent	A configured Copilot experience that may combine instructions, knowledge, tools, channels, identity, actions, memory, and triggers.

1.4 Assumptions That Must Be Validated

- The bank is using an eligible commercial Microsoft 365 tenant and signs users in with Microsoft Entra work accounts.
- The bank has a negotiated or accepted Microsoft DPA and Product Terms and understands which licenses and add-ons provide Purview, SharePoint, audit, eDiscovery, and Copilot controls.
- The bank can identify which Copilot UI, agent, model provider, knowledge source, and channel produced each material result.
- The bank has authority to process each data class for the approved purpose and has addressed privacy notices, contractual restrictions, records requirements, and customer commitments.
- The bank will not treat a generated answer as authoritative evidence without source review, especially for credit, fraud, BSA/AML, legal, accounting, HR, cybersecurity, or customer communications.

2. Product Boundaries and Why They Matter

The same word 'Copilot' can describe products with different grounding, contractual roles, storage, controls, and user identities. The safest bank control is a product allowlist tied to Entra identity and device policy, not a training instruction that asks employees to identify the right surface by visual appearance alone. [2] [5] [6] [29]

Surface	Typical data access	Protection boundary	Bank risk focus

Surface	Typical data access	Protection boundary	Bank risk focus
Licensed Microsoft Copilot for work	Microsoft Graph work data, open files, chats, meetings, mail, apps, web, and approved agents, subject to product and license capabilities.	Microsoft DPA, Product Terms, enterprise data protection, tenant identity and permissions.	Oversharing, source accuracy, labels, retention, web search, model provider, and decision use.
Copilot Chat with work account	Usually web-grounded, with bank data introduced through upload, paste, Outlook scope, open content, or agents.	Enterprise data protection for eligible work use, plus a separate Bing web-search processing boundary.	Users may assume it cannot see bank data. Uploaded, pasted, or scoped content disproves that assumption.
Copilot Studio agent	Configured knowledge, tools, connectors, Dataverse, SharePoint, HTTP, channels, triggers, and credentials.	Microsoft and Power Platform controls plus every configured source, channel, provider, and action boundary.	Authentication, maker credentials, external audience, prompt injection, actions, transcripts, source permissions, and environment governance.
Consumer Copilot or personal account	Consumer conversation and uploaded content under consumer settings and privacy controls.	Consumer terms and controls, not the bank's enterprise DPA and tenant governance.	Bank data should be prohibited. Training and history choices differ from work use.
Third-party or preview model inside Copilot	Selected input and output required for the feature, with handling based on provider role and feature terms.	May be Microsoft-hosted, a Microsoft subprocessor, or an independent processor.	Default enablement, retention, geography, certifications, legal terms, monitoring, and data-class eligibility.

2.1 Work Account Versus Consumer Account

A bank should regard the work-account indicator and enterprise-data-protection status as a security control. Consumer Copilot documentation includes user choices about whether conversations may improve models and describes consumer conversation history. Those controls are not a substitute for Microsoft's enterprise commitments. A personal account, a browser profile signed into a consumer identity, or an unapproved mobile app should therefore be treated as an unapproved external AI service. [2] [29] [30]

Policy test

If the bank cannot centrally prove the identity, product surface, model-provider state, audit source, and governing terms for a session, the session should not receive bank confidential information.

2.2 Copilot Chat Can Still Receive Organizational Data

Copilot Chat is often described as web-grounded and not tenant-wide work-data grounded. That description can be misunderstood. Organizational data can enter when a user uploads or pastes a file, invokes Copilot in Outlook, works with open content, or uses an agent. The bank should label these as different modes in training and control documentation. [5] [6]

3. End-to-End Data Lifecycle

A useful risk assessment follows the data from user input through grounding, model inference, response display, storage, audit, and deletion. Each step has a different control owner and different failure modes. [1] [3] [4]

Stage	What happens
1. Identity and session	The user signs in. Entra identity, license, group assignment, Conditional Access, device posture, and the selected Copilot surface determine the available controls.
2. User and system input	The effective prompt can include typed text, pasted content, uploaded files, chat history, agent instructions, trigger payloads, and context from the current application.
3. Grounding	The orchestrator may query Microsoft Graph, SharePoint, OneDrive, mail, Teams, approved connectors, agent knowledge, or the public web. Permissions and DLP may filter content.
4. Model inference	The assembled request is sent to a Microsoft-hosted model or an allowed model provider. Geographic, subprocessor, retention, and certification treatment depends on the provider path.
5. Response and citations	The model produces an answer. Citations may point to source content. Output can contain inaccurate statements, sensitive facts, derived classifications, or more information than the user intended to reveal.
6. Storage and compliance copy	Prompts, responses, citations, and interaction metadata can be stored in Microsoft 365 services, including Exchange-backed locations, and may be subject to retention, holds, audit, eDiscovery, and compliance review.
7. Downstream use	The user may copy output into mail, documents, tickets, source code, decisions, customer communications, or another system. The receiving system becomes part of the record and access boundary.
8. Disposition and evidence	Retention policies, holds, user deletion, eDiscovery purge, source-system deletion, agent transcript settings, and vendor commitments determine when copies are deleted and what evidence remains.

3.1 Prompt Composition Is Broader Than Typed Text

Risk reviews that inspect only the visible text box are incomplete. A model request can include system instructions, prior turns, retrieved work data, document excerpts, file metadata, agent topics, connector results, trigger payloads, and generated web queries. The bank should document the full effective input for each approved use case and should test it with audit records and synthetic data. [1] [3] [5] [39]

3.2 Grounding Does Not Copy the Whole Tenant

Microsoft describes an orchestration process that retrieves relevant information subject to the user's permissions. Copilot does not need to export the entire tenant to answer a question. It can still process sensitive excerpts, names, numbers, and metadata selected for a specific response. Data minimization therefore depends on source permissions, retrieval relevance, prompt design, and policy enforcement, not merely on the absence of a bulk export. [1] [3] [4]

3.3 Response Content Can Be Newly Sensitive

A response may combine individually innocuous facts into a sensitive conclusion, summarize a confidential investigation, infer a customer condition, or expose a broad pattern across multiple sources. Classification should follow the content of the response, not only the labels of the inputs. Microsoft documents label inheritance behavior for supported labeled content, but banks should test complex multi-source prompts and downstream copies. [4] [13]

3.4 User Deletion Is Not the Same as Immediate Destruction

When interaction content is subject to retention or legal hold, deleting it from a user interface does not necessarily remove the compliance copy. Microsoft describes hidden Exchange storage and hold locations that preserve content until the retention and disposition conditions are satisfied. Legal, privacy, records, and incident teams need a documented deletion path before deployment. [9] [11]

4. Security, Privacy, and Contractual Commitments

4.1 Foundation-Model Training

Microsoft states that prompts, responses, and data accessed through Microsoft Graph are not used to train foundation models for enterprise Copilot and Copilot Chat. Optional user feedback can be used to improve Copilot services, but Microsoft distinguishes this from training foundation models and provides administrative controls over feedback. The bank should preserve the exact contractual and product language in its due-diligence file and recheck it when model providers change. [1] [2]

Do not compress this into 'Microsoft does not use our data.'

The accurate statement is narrower: defined enterprise prompts, responses, and Graph data are not used to train foundation models. The service still processes the data to produce responses, stores interaction content, generates audit and compliance data, may create Bing search queries, may involve configured subprocessors, and may process optional feedback. [1] [2] [5] [7] [9] [23] [24]

4.2 Human Access and Abuse Monitoring

Microsoft documents that Copilot service use of Azure OpenAI is opted out of abuse-monitoring processes that include human review. That reduces one data-exposure path, but it should not be interpreted as a promise that no human can ever access any related information. Authorized support, security, legal, compliance, and service operations can involve controlled access under applicable contracts, access policies, and features such as Customer Lockbox. The bank should validate support-access terms, logging, approvals, and emergency exceptions. [1] [15]

4.3 Encryption and Tenant Isolation

Microsoft states that enterprise Copilot uses Microsoft 365 protections including encryption in transit and at rest, tenant isolation, identity controls, permissions, and compliance tooling. These are strong

platform controls. They do not prevent an authorized but overprivileged user from retrieving data, a user from placing data in the wrong prompt, an agent from using maker credentials, or an output from being copied to an unprotected location. [2] [3] [4]

4.4 Sensitivity Labels and Encryption Rights

For supported protected content, Copilot respects sensitivity labels and encryption usage rights. Microsoft notes that a user generally needs both VIEW and EXTRACT rights for encrypted content to be used in a Copilot response. This makes rights design material. A bank should not assume that a label name alone blocks processing; the label's encryption and rights configuration, DLP policy, source support, and response behavior must be tested. [4] [12] [13]

4.5 Customer Key and Customer Lockbox

Customer Key can extend customer-controlled key protection to Microsoft Copilot interactions at rest for supported licensing and workloads. Customer Lockbox can require customer approval for covered Microsoft support access to Microsoft 365 content, including Copilot interaction content stored through Exchange. These controls can improve a defense-in-depth design, but they require licensing, operational ownership, recovery planning, and documented exception handling. [14] [15]

4.6 Accuracy, Reliability, and Human Judgment

Microsoft warns that generated content can be inaccurate and should not replace human judgment. In a bank, this is an integrity and conduct risk as much as a quality issue. A polished answer may contain invented policy language, obsolete regulatory citations, incorrect account facts, fabricated meeting decisions, or incomplete adverse-action reasons. Material outputs need traceable sources, qualified review, and a prohibition on automatic reliance unless the use case has a tested control framework. [1] [51] [52] [55]

5. Permissions, Oversharing, and Connectors

Copilot's Microsoft Graph access is permission-aware: it should retrieve only content the user can access. That is good design, but it turns permission quality into the central confidentiality control. Broad group membership, company-wide links, stale sharing links, inherited access, external sharing, and incorrectly mapped connector permissions can all become visible through natural-language retrieval. [1] [3] [20] [22]

5.1 Why Copilot Changes the Consequence of Oversharing

Before Copilot, a user might need to know that a file exists, search for the right keywords, open multiple documents, and interpret them. Copilot lowers those friction costs. It can locate, summarize, and combine content across sources. The underlying authorization may be unchanged, but the probability and scale of discovery rise. This is an amplification of existing access, not necessarily a new access grant. [3] [17] [18]

5.2 Temporary Search Restrictions Versus Real Access Controls

Restricted SharePoint Search and Restricted Content Discovery can reduce organization-wide search and Copilot discovery while the bank remediates permissions. Microsoft explicitly describes these as temporary or limited controls, and Restricted SharePoint Search is not a security boundary. Restricted access control, group-based site access, corrected sharing, and least-privilege permissions are stronger long-term controls. [16] [17] [21]

5.3 Connector Permission Mapping

Copilot connectors can index line-of-business content. If the connector maps source access to 'Everyone,' omits an access-control list, or cannot preserve a source's permission model, indexed content can be retrievable by more users than intended. Due diligence must cover connector schema, identity mapping, delta updates, revocation latency, deleted records, external users, nested groups, and failure behavior. Permission accuracy should be tested with positive and negative user personas before production. [22]

5.4 Oversharing Remediation Sequence

1. Inventory high-value SharePoint sites, OneDrive locations, Teams, mailboxes, and connectors that contain customer NPI, legal, fraud, HR, finance, cybersecurity, or strategic information. [18] [20]
2. Run broad-permission and sharing-link reports. Prioritize sites with large audiences, external users, anonymous links, dormant owners, or sensitive labels. [19] [20]
3. Apply temporary discovery restrictions where needed, but assign owners and dates for actual permission remediation. [16] [17]
4. Use restricted access control or equivalent identity-based boundaries for the most sensitive sites. [21]
5. Test with synthetic canary files and personas who should and should not retrieve the content. Repeat after group, connector, and agent changes. [22]

6. Web Grounding and External Data Flows

Web grounding improves currency and breadth but creates a separate data flow. Microsoft says Copilot generates a short search query based on the prompt, rather than sending the full prompt to Bing. It also describes removal of user and tenant identifiers and limits on advertising, behavioral tracking, improvement of Bing, and foundation-model training. The query can still contain sensitive terms derived from the prompt, uploaded content, or visible context if those terms are relevant to the search. [2] [5] [7]

6.1 Legal and Residency Boundary

Microsoft describes Bing as a separate service and a different legal role for generated search queries. The Microsoft 365 DPA and some Microsoft 365 geographic commitments do not apply to the Bing query in the same way they apply to stored Copilot interaction content. For bank analysis, this is not a semantic detail. It affects the data-flow inventory, privacy notice analysis, contract review, residency statement, subprocessor record, and DLP decision. [2] [5] [7] [42]

6.2 Recommended Web-Search Rule

Default rule

Allow web search for public and ordinary internal research. Block web search when prompts contain customer NPI, account identifiers, employee restricted data, privileged legal matter names, security incident details, unreleased financial information, SAR-related content, authentication material, or payment account data. Use Purview DLP and product controls, then test that the intended search restriction actually takes effect. [7] [12] [42] [50] [58]

6.3 Web Search Does Not Cure Hallucination

Search-grounded responses may cite low-quality, stale, promotional, or contextually wrong sources. Citations demonstrate retrieval, not correctness. Bank users should prefer primary sources, open and read the cited material, verify dates and jurisdiction, and retain evidence for any material compliance or customer decision. [1] [55]

7. Model Providers and 2026 Subprocessor Changes

Microsoft now documents three distinct provider roles: Microsoft-hosted models, model providers acting as Microsoft's subprocessors, and independent processors operating under their own terms. The role determines contract, data handling, geography, retention, and control implications. A bank should manage model providers as configuration-controlled third parties, not as an invisible technical implementation detail. [23] [24] [25] [26] [27]

7.1 OpenAI-Operated Models

Microsoft's 2026 documentation distinguishes OpenAI-operated models from models operated by Microsoft through Azure OpenAI. OpenAI was added as a Microsoft Copilot subprocessor, and Microsoft states that eligible commercial users may have OpenAI-operated models enabled unless an administrator disables them. Microsoft says it uses the OpenAI Responses API with Zero Data Retention, subject to feature limitations, and applies the Microsoft DPA, Product Terms, and enterprise data protection to this subprocessor path. [23] [25] [26]

The same documentation identifies material assurance differences. OpenAI-operated models are excluded from some in-country processing commitments and do not carry specified Microsoft service attestations, including FedRAMP High, a PCI DSS Attestation of Compliance, HITRUST documentation, and SOC 1 Type 2 for that provider path. A bank should map those exclusions to the data and use case before leaving the provider enabled. [23]

7.2 Anthropic-Operated Models

Microsoft documents Anthropic models under more than one path. A standard subprocessor path can be governed under the Microsoft DPA and Product Terms, with administrative controls and regional exclusions. Selected preview models with data retention can operate as an independent processor

under Anthropic terms. Microsoft documents provider-side retention for those features, including longer retention for certain policy or safety events. These preview paths should be default-denied for bank data until legal, privacy, security, retention, and examination concerns are resolved. [24] [25] [27]

7.3 Provider Control Standard

Control question	Required evidence
Which provider processed the request?	Admin configuration, product telemetry, feature documentation, and an auditable provider inventory.
Was the provider enabled by default?	Dated screenshots or exports of tenant settings, eligible users and groups, and change tickets.
What legal role applies?	Microsoft-hosted, Microsoft subprocessor, or independent processor, with the exact DPA, Product Terms, and provider terms archived.
What is retained and for how long?	Feature-specific retention, Zero Data Retention limitations, safety retention, logs, feedback, and support records.
Where can processing occur?	Data-at-rest and in-process commitments, EU Data Boundary treatment, in-country exclusions, and disaster-recovery locations.
Which assurance reports apply?	SOC, ISO, PCI, FedRAMP, HITRUST, penetration testing, and bridge letters scoped to the actual provider path.
How is access disabled?	Group-level controls, emergency disable process, license effect, and evidence that existing conversations or agents stop using the provider.

8. Purview, DLP, Audit, Retention, and Residency

8.1 Prompt and Response Storage

Microsoft stores Copilot interaction content, including prompts and responses, in Microsoft 365 services and describes Exchange-backed hidden folders for retention and eDiscovery. Interaction history can be visible to users, while compliance copies may remain after user deletion when retention or holds apply. The bank should treat prompts as records that may contain customer information, employee information, investigations, legal strategy, and security data. [1] [5] [8] [9] [11]

8.2 Audit and Compliance Access

Purview audit can record Copilot activity and references to files or sensitivity labels. DSPM for AI and eDiscovery can expose interaction content to authorized compliance roles. Communication Compliance can analyze prompts and responses for risky or confidential content. These controls support oversight but create a privileged monitoring function. Assign narrow roles, use pseudonymization where supported, monitor searches and exports, separate investigators from platform administrators, and document employee-notice and labor requirements. [10] [11]

8.3 DLP Capabilities

DLP control	Value	Important limitation
Block external web search for sensitive prompt text	Keeps specified sensitive terms from becoming generated Bing queries.	Depends on supported sensitive-information detection and policy propagation. Validate with test prompts.
Block processing of sensitive prompts	Can stop Copilot from handling selected sensitive prompt content.	Microsoft documents preview or rollout status for parts of this capability. Confirm tenant availability.
Block labeled files and emails as grounding	Prevents selected labels from being processed by Copilot.	A blocked item can still appear as a citation in some cases. Validate response and citation behavior.
Block external email grounding	Reduces prompt-injection and data-contamination paths from outside senders.	Coverage and rollout status must be confirmed. Calendar invitations are not supported.
Sensitive information types in typed prompt	Detects defined NPI, payment, identity, and other patterns in user text.	Directly uploaded files are not content-scanned by this Copilot DLP location. An unlabeled sensitive upload can bypass this inspection path.
Policy rollout	Central control across scoped users.	Microsoft documents that changes can take up to four hours. Emergency controls need additional layers.

The direct-upload limitation is a material bank control gap. If a user uploads an unlabeled file containing account numbers or other NPI, the Copilot DLP location does not content-scan the file in the same way it scans typed prompt text. Compensating controls include endpoint and browser controls, mandatory labeling at rest, restricted upload features where available, user training, activity monitoring, and an explicit data-class rule. [12] [13]

8.4 Retention Design

Retention should be based on record purpose and legal requirements, not on a blanket desire to keep everything for future AI value. The bank should decide whether general productivity interactions are transitory, whether certain business interactions are records, and how holds apply. Shorter retention can reduce exposure, but retention policies, holds, investigations, and user deletion behavior must be reconciled. A zero-day retention policy is not automatically immediate erasure if holds or processing queues apply. [9] [11] [38]

8.5 Data Residency

Microsoft documents data-at-rest commitments for Copilot interaction content and semantic index data under Microsoft 365 residency programs. It also states that model calls may be processed in the United States, European Union, or other regions, depending on the service and provider. Web-search queries and some subprocessors have separate exclusions. A bank data-residency statement should therefore distinguish data at rest, transient processing, support access, logs, web queries, subprocessors, independent processors, and disaster recovery. [1] [5] [8] [23] [24]

9. Copilot Studio Agents and Automation

Copilot Studio changes Copilot from a general assistant into a configurable application and automation platform. An agent may combine instructions, knowledge, tools, connectors, authentication, channels, triggers, analytics, transcripts, and model providers. The control boundary is the entire agent design and deployment, not the chat window. [31] [32] [41]

9.1 Authentication and Audience

New agents can default to Microsoft Entra authentication, but maker choices and tenant policy determine whether no-authentication publishing is possible. A link-accessible or external-channel agent can expose content beyond the bank if the knowledge and action design does not enforce user identity. Power Platform data policies should block no-authentication modes and unapproved channels for bank environments. [31] [32] [33]

9.2 Maker Credentials Versus End-User Credentials

The credential mode is one of the most important agent controls. With end-user credentials, a connector or action can align access with the interacting user's permissions. With maker or author credentials, every allowed user may be able to cause the agent to access data or execute actions using the maker's authority. Microsoft documents that both credential patterns can be available by default in some configurations. Banks should enforce end-user credentials for interactive agents unless a documented service-identity design is approved. [33] [34]

Autonomous-agent conflict

Scheduled, event-driven, and autonomous agents often cannot rely on an interactive user's credentials. If the bank permits them, use a dedicated least-privilege service identity, narrow connector scopes, transaction limits, approval gates, idempotency, kill switches, and reconciliation. Do not use a human maker's broad account as an unattended service identity. [33] [34] [41]

9.3 Knowledge Sources

SharePoint, Dataverse, and supported connectors can use the agent user's Entra identity to trim results. Uploaded document knowledge is stored in Dataverse and is not protected by a per-user source permission model in the same way as SharePoint. Treat the agent audience and environment permissions as the access boundary for uploaded knowledge. Do not upload broad customer, legal, HR, fraud, cybersecurity, or SAR repositories into a generally available agent. [35]

9.4 Actions, Connectors, and Prompt Injection

An agent that can read email, external websites, documents, tickets, or messages can ingest malicious instructions embedded in content. If the same agent can send mail, change records, transfer data, or operate a user interface, prompt injection can become an authorization problem. Separate read and write tools, treat retrieved content as untrusted data, use allowlisted parameters, require confirmation

for material actions, and validate the final transaction against policy outside the model. [32] [40] [41] [55]

9.5 Transcripts, Historical Activity, and Audit

Copilot Studio can store conversation transcripts in Dataverse, with Microsoft documenting a default retention period for transcript records that can be configured. Agent activity may also be managed through Microsoft 365 compliance services, and audit events can include thread identifiers and related metadata. Historical activity can record user input, trigger payloads, responses, and, for supported models, reasoning information. Retention and access must be designed across both Dataverse and Microsoft 365 locations. [36] [37] [38] [39]

9.6 Computer Use

Computer-use agents can interact with application user interfaces and may operate using maker-provided credentials on a configured machine. Monitoring can include screenshots and tool logs, which may capture bank data, credentials displayed on screen, customer records, or regulated workflows. This should be treated as privileged robotic-process automation with AI uncertainty, not as ordinary chat. Require isolated machines, nonhuman identities, minimal application roles, session controls, screenshot retention rules, transaction limits, and independent reconciliation. [40] [44]

9.7 Minimum Agent Approval Packet

- Named business owner, technical owner, data owner, risk tier, approved purpose, prohibited uses, and sunset date.
- Architecture showing environment, model provider, instructions, knowledge, connectors, actions, identity, channels, network paths, transcripts, logs, and downstream systems.
- Data inventory by class and record type, including prompts, responses, retrieved excerpts, files, metadata, logs, screenshots, and error payloads.
- Positive and negative permission tests for every knowledge source and connector, including external users, terminated users, group changes, and deleted records.
- Prompt-injection, data-exfiltration, unsafe-action, hallucination, denial-of-service, and business-continuity tests.
- Human-approval gates, monetary or operational limits, kill switch, rollback plan, incident contacts, audit evidence, and periodic recertification.

10. Bank Regulatory and Compliance Analysis

The following mapping is a risk-management aid, not a legal conclusion. Applicability depends on charter, regulator, asset size, state law, data, customer relationship, contract, and use case. The central principle is consistent across authorities: outsourcing processing does not outsource accountability.

10.1 GLBA Information Security and Interagency Guidelines

The Interagency Guidelines require a comprehensive written information-security program appropriate to the bank's size, complexity, and activities. The program must protect the security and confidentiality

of customer information, address foreseeable threats, prevent unauthorized access, assess risk, use appropriate controls, train staff, test controls, oversee service providers, adjust to change, and report to the board. Copilot processing of customer NPI belongs inside this program, including interaction content and agent logs maintained on the bank's behalf. [46]

The service-provider provisions require due diligence, contractual safeguards, and risk-based monitoring using audits, test summaries, or equivalent evaluations. The bank should be able to show how Microsoft's DPA, Product Terms, subprocessor commitments, incident terms, audit reports, data locations, support access, deletion, and configuration controls meet the bank's own risk assessment. [46]

10.2 Third-Party Risk Management

The 2023 interagency guidance applies a lifecycle of planning, due diligence and selection, contract negotiation, ongoing monitoring, and termination. Oversight should be commensurate with risk and criticality. For Copilot, the relationship includes Microsoft, relevant model providers, Copilot Studio connectors, agent publishers, implementation partners, and possibly downstream data sources or action services. The bank should maintain a complete inventory and reassess material configuration changes as third-party changes. [47]

10.3 Cloud Shared Responsibility

FFIEC cloud guidance emphasizes shared responsibility and warns institutions not to assume that a cloud provider has implemented every required control. Microsoft controls infrastructure and many service layers. The bank controls identity, licenses, permissions, labels, DLP, provider enablement, SharePoint governance, agent design, user behavior, retention choices, and output use. The risk assessment should assign each control to an owner and preserve evidence of operation. [48]

10.4 Regulation P and Purpose Limitation

Regulation P governs privacy notices and limits on disclosure of nonpublic personal information, subject to service-provider, transaction-processing, consent, and other exceptions. The bank should document why Microsoft and any configured model or connector may receive each data class, whether the disclosure fits the bank's privacy notice and legal exceptions, and whether the data is used only for the approved purpose. An independent-processor preview model may require a different analysis from a Microsoft subprocessor. [48] [23] [24] [25]

10.5 SAR Confidentiality and BSA/AML

A SAR and information that would reveal the existence of a SAR are confidential under 31 CFR 1020.320. A general-purpose Copilot prompt can be retained, indexed for compliance, displayed to administrators, sent through web grounding, or processed by a configured provider. The safest default is to prohibit SARs and SAR-revealing information in general Copilot use. Any specialized BSA/AML use should be isolated, legally reviewed, narrowly accessed, web-disabled, provider-restricted, retention-controlled, and tested for unauthorized disclosure. [50]

10.6 Fair Lending, ECOA, and Adverse Action

CFPB circulars state that ECOA and Regulation B apply regardless of algorithmic complexity and require accurate, specific principal reasons for adverse action. Copilot should not invent, select, rewrite, or substitute adverse-action reasons unless the reasons come from the bank's validated decision system and are verified. A language model's plausible explanation is not evidence of the actual factors used. Fair-lending testing must address differential error, tone, recommendations, and assistance across protected groups and proxies. [51] [52]

10.7 Model Risk Management Nuance

The April 2026 interagency model-risk guidance replaces SR 11-7 and SR 21-8 and is expected to be most relevant to banking organizations over $30 billion. Its definition centers on complex quantitative methods that produce estimates. A footnote expressly states that generative AI and agentic AI models are outside the guidance's scope because they are novel and evolving. It also states that a bank's risk-management and governance practices should guide controls for tools outside scope. [53]

The practical conclusion is not that Copilot is ungoverned and not that full model validation automatically applies. The bank should place Copilot in its AI, technology, third-party, information-security, operational, and compliance governance. If Copilot develops, changes, explains, or feeds a traditional quantitative model, or if a downstream decision model relies on its output, the in-scope model and the interface may require model-risk review. Risk-based testing, inventory, documentation, effective challenge, monitoring, and use limitations remain sound practices even when the generative component is outside formal guidance scope. [53] [54] [55]

10.8 Records, Litigation, Investigations, and Employee Monitoring

Stored prompts and responses may become business records, discoverable material, investigative evidence, or employee-monitoring data. Retention must align with record schedules, legal holds, privacy notices, labor requirements, internal investigations, and secure deletion. Broad administrative access to prompt text can expose privileged or need-to-know information. Legal, records, HR, privacy, and audit teams should approve role design and review procedures. [9] [10] [11]

10.9 Cyber Incident Notification

Qualifying computer-security incidents can trigger rapid regulator notification, including the 36-hour rule for OCC-supervised banks and parallel federal banking rules. Covered bank service providers must notify affected bank customers of certain material disruptions. Copilot and agent incident plans should define how Microsoft or another provider notifies the bank, how the bank obtains scope and customer-impact evidence, and who decides whether regulatory, law-enforcement, SAR, customer, state, contractual, or insurer notice is required. [46] [56]

10.10 Bank Service Company Act

Covered service relationships can require written notice to the appropriate federal banking agency under the Bank Service Company Act. The bank should ask regulatory counsel or its primary regulator whether the particular Microsoft service, Copilot Studio action, or outsourced bank function changes an

existing notice or requires a new one. Do not assume that a preexisting Microsoft 365 relationship automatically covers every AI-enabled service activity. [57]

10.11 Payment Account Data and PCI DSS

PCI DSS applies to environments that store, process, or transmit payment account data. Entering cardholder data into Copilot can expand scope and create copies in prompts, responses, logs, transcripts, citations, screenshots, or support records. Microsoft documentation notes that some OpenAI-operated model paths do not carry a PCI DSS Attestation of Compliance. The conservative default is to prohibit full payment-card data and sensitive authentication data in Copilot, use tokens or redaction for approved analysis, and involve the bank's PCI assessor before any exception. [23] [58] [59]

10.12 Regulatory Mapping Table

Authority or obligation	Copilot implication	Evidence the bank should retain
GLBA and Interagency Information Security Guidelines [46]	Customer NPI in prompts, grounding, output, or logs is customer information handled by or on behalf of the bank.	Risk assessment, board reporting, safeguards, staff training, tests, provider contract, audit reports, incident and deletion procedures.
Third-party lifecycle guidance [47]	Microsoft, model providers, connectors, publishers, and integrators can be part of one service chain.	Inventory, criticality, due diligence, contract analysis, ongoing monitoring, concentration and exit plans.
FFIEC cloud statement [48]	The bank owns configuration, identity, data, use, and monitoring responsibilities even when Microsoft owns infrastructure controls.	Shared-responsibility matrix, configuration baseline, control owners, tests, independent review.
Regulation P [48]	NPI processing must fit notice, purpose, exception, and third-party sharing requirements.	Data-flow map, legal basis, privacy notice mapping, provider role, data minimization and deletion.
SAR confidentiality [50]	Prompts, summaries, or questions can reveal that a SAR exists.	Default prohibition, specialized exception design, access list, web and provider controls, monitoring.
ECOA and Regulation B [51] [52]	AI cannot excuse inaccurate or nonspecific reasons for credit adverse action.	Actual decision factors, source traceability, reviewer evidence, fair-lending testing, approved customer language.
Revised model-risk guidance [53]	Generative and agentic AI are outside formal scope, but governance should guide controls. Downstream quantitative models can still be in scope.	AI inventory, risk tier, testing, use limits, monitoring, decision lineage, model-risk determination.
Incident notification [56]	Material provider or agent incidents can require rapid escalation and bank notification.	Contract notice terms, contact list, severity mapping, decision log, provider evidence, tabletop results.
PCI DSS [58]	Payment account data in Copilot can expand the cardholder-data environment and create uncontrolled copies.	Prohibition or tokenization, data-flow evidence, provider AOC scope, assessor determination, monitoring.

11. Consolidated Risk Register

Ratings assume use before controls and a target after the baseline in Section 13. Each bank should adjust probability and impact for its own data, architecture, use cases, regulatory profile, and control maturity.

Risk scenario	Why it matters to a bank	Inherent	Primary controls	Target
Existing SharePoint or OneDrive oversharing	Users can retrieve and combine data they technically can access but were not expected to find.	High	Permission reports, owner review, restricted access control, link cleanup, canary tests, quarterly recertification.	Medium
Wrong identity or consumer Copilot	Bank data leaves the enterprise contract, audit, identity, and policy boundary.	High	Block personal accounts where feasible, managed browser and device, allowlist, UI training, monitoring, sanctions.	Low
Sensitive typed prompt	NPI, HR, legal, incident, or authentication details enter a retained interaction record.	High	Data-class rules, prompt DLP, labels, training, web-search block, audit alerts, minimization.	Medium
Sensitive direct file upload	Unlabeled files are not content-scanned by the Copilot DLP location, creating a coverage gap.	High	Mandatory labeling, upload restrictions, endpoint controls, approved repositories, activity review, user policy.	Medium
Generated Bing query	Sensitive terms can cross a separate service and legal boundary with different residency treatment.	High	Disable web for sensitive classes, DLP web-search block, test query behavior, separate data-flow record.	Low
Model-provider change or default enablement	Data reaches an unassessed provider with different geography, retention, or attestations.	High	Provider default deny, group allowlist, change alerts, quarterly settings export, contract and subprocessor review.	Low
Hallucinated policy or regulation	Staff act on false or obsolete information, creating compliance, conduct, or legal risk.	High	Primary-source citations, qualified review, prohibited-use list, retrieval tests, no automatic reliance.	Medium
Incorrect customer communication	Generated content misstates fees, rights, reasons, terms, privacy, or complaint handling.	High	Approved templates, fact sources, reviewer sign-off, output constraints, sampling and complaint monitoring.	Low
Prompt injection through retrieved content	Malicious instructions cause disclosure, tool misuse, or unsafe action.	High	Treat content as untrusted, separate read/write, action allowlists, confirmation, content controls, red-team tests.	Medium
Maker credential privilege escalation	Agent users cause actions or retrieval using the author's broader authority.	High	End-user credentials, dedicated service identities, minimal scopes, no human maker for unattended work, action limits.	Low
Unauthenticated agent or external channel	Unknown users access knowledge or trigger actions.	High	Entra authentication, block no-auth mode, channel allowlist, external-user test, environment DLP.	Low
Autonomous or scheduled action error	An incorrect action executes without contemporaneous human review.	High	Approval gates, value limits, dual control, idempotency, reconciliation, kill switch, rollback, alerts.	Medium
Computer-use agent captures secrets	Screenshots, logs, or UI automation expose credentials and customer data.	High	Isolated machine, nonhuman identity, masked screens, restricted apps, screenshot retention, session monitoring.	Medium

Risk scenario	Why it matters to a bank	Inherent	Primary controls	Target
Connector permission-mapping error	Indexed business data is visible to unintended users.	High	Source ACL fidelity, negative persona tests, revoke and delete tests, connector monitoring, fail closed.	Low
Interaction monitoring abuse	Privileged administrators view sensitive employee, legal, customer, or investigation prompts.	High	Least privilege, separation of duties, pseudonymization, search logging, dual approval for export, periodic audit.	Low
Retention mismatch	The bank keeps prompts too long, deletes records too soon, or cannot execute legal holds and privacy deletion.	High	Record classification, retention policy, hold and purge tests, Dataverse and Exchange alignment, owner approval.	Medium
Regional processing mismatch	Actual provider or web-search processing conflicts with commitments or bank policy.	Medium	Data-flow map by provider, residency review, excluded-data classes, contract record, quarterly validation.	Low
License or feature gap	The bank assumes a Purview, audit, DLP, or SharePoint control that its license or rollout does not provide.	High	Entitlement inventory, tenant test, release-state evidence, compensating controls, deployment gate.	Low
Vendor or feature change	Rapid product changes invalidate controls, training, or risk conclusions.	High	Message Center and roadmap monitoring, subprocessor alerts, quarterly reassessment, change freeze for high-risk agents.	Medium
Concentration and exit risk	Dependence on Microsoft identity, files, productivity, AI, and agents creates broad operational impact.	Medium	Business continuity, manual fallback, exportability, agent inventory, contract exit assistance, outage tabletop.	Medium

12. Data Classification and Use Decision Matrix

The matrix is intentionally conservative. 'Conditional' means the use case, product surface, identity, provider, web-search state, data source, retention, and controls are approved. It does not mean every employee may enter that data.

Bank data class	Work Copilot	Copilot Chat	Approved Studio agent	Consumer or unapproved AI
Public and approved external content	Allowed	Allowed	Allowed	Allowed only if bank policy permits the service and no internal context is added
Internal, low-sensitivity operational content	Allowed with work account	Allowed with work account	Conditional	Prohibited
Confidential business information	Conditional	Conditional; disable web when terms are sensitive	Conditional; named audience and source controls	Prohibited

Bank data class	Work Copilot	Copilot Chat	Approved Studio agent	Consumer or unapproved AI
Customer NPI and account information	Conditional exception with minimization, labels, DLP, retention, and approved purpose	Conditional exception; prefer no upload and no web	High-risk approval with strict identity and source controls	Prohibited
Employee restricted, legal, investigations, fraud, or cybersecurity	High-risk exception	High-risk exception; web off	High-risk isolated agent	Prohibited
Payment-card account data	Tokenized or redacted only unless PCI-approved architecture	Tokenized or redacted only	Only in PCI-approved scope	Prohibited
Authentication secrets, private keys, passwords, access tokens	Prohibited	Prohibited	Use managed secret stores and connector references, never prompt content	Prohibited
SAR or information revealing SAR existence	Prohibited by default; specialized legal exception only	Prohibited	Specialized isolated design only	Prohibited
Material credit, BSA/AML, legal, accounting, or customer decision	Drafting or analysis only with qualified human decision	Research support only	No autonomous final decision without separate approval	Prohibited

12.1 Use-Case Tiers

Tier	Examples	Approval
Tier 1: Low	Summarize public information, draft internal agendas, rewrite non-sensitive text, generate ideas.	Business owner and standard controls.
Tier 2: Moderate	Summarize internal policies, prepare meeting notes, search ordinary internal content, draft employee communications.	Data owner, security baseline, retention, user training, and sampling.
Tier 3: High	Customer NPI, legal, HR, cybersecurity, finance, fraud, external communications, action agents, sensitive connectors.	Formal risk assessment, legal and compliance review, security testing, executive owner, periodic recertification.
Tier 4: Restricted	SAR content, secrets, full card data, autonomous account changes, final credit or BSA decisions, computer use in critical systems.	Prohibited unless a specialized architecture receives senior exception approval and regulator or counsel input as appropriate.

13. Required Control Baseline

The following baseline should be implemented and evidenced before broad deployment. High-risk use cases require additional controls.

13.1 Governance and Inventory

- Name an accountable executive, AI risk committee, platform owner, data owners, privacy, compliance, legal, security, records, audit, model-risk liaison, and business owners. [47] [53] [54] [55]
- Maintain an inventory of Copilot licenses, eligible users, model providers, agents, connectors, knowledge sources, channels, credential modes, environments, data classes, retention, and use cases. [47] [53] [54] [55]
- Adopt a risk-tiering method with deployment gates, prohibited uses, exception approval, change control, periodic recertification, and retirement requirements. [47] [53] [54] [55]
- Treat model-provider enablement, web search, agent publication, connector scope, and maker-credential use as material configuration changes. [47] [53] [54] [55]

13.2 Identity, Device, and Session

- Require Microsoft Entra work accounts, phishing-resistant MFA where feasible, managed devices for sensitive use, Conditional Access, and session controls. [45]
- Restrict personal accounts and unapproved AI sites with identity, browser, network, endpoint, and policy controls. Train users to verify enterprise protection before entering data. [45]
- Use separate admin accounts, privileged identity management, just-in-time access, break-glass procedures, and audit for Copilot, Purview, SharePoint, and Power Platform administration. [45]

13.3 Data and Permissions

- Complete SharePoint, OneDrive, Teams, mailbox, and connector oversharing remediation before broad graph-grounded deployment. [4] [13] [16] [17] [20] [21] [22]
- Apply sensitivity labels and encryption to customer NPI and other restricted content. Validate VIEW and EXTRACT rights and response label behavior. [4] [13] [16] [17] [20] [21] [22]
- Use restricted access control for high-sensitivity sites. Use temporary discovery restrictions only with a remediation owner and date. [4] [13] [16] [17] [20] [21] [22]
- Minimize data in prompts and test downstream output classification and copy behavior. [4] [13] [16] [17] [20] [21] [22]

13.4 DLP, Web, and Provider Controls

- Implement Copilot DLP policies for sensitive prompt text, selected labels, web search, and external-email grounding where supported. [7] [12] [23] [24] [27]
- Compensate for direct-upload content-inspection gaps with mandatory labeling, approved repositories, endpoint control, monitoring, and policy. [7] [12] [23] [24] [27]
- Default-deny preview and independent-processor models. Allowlist provider paths by user group and data class after review. [7] [12] [23] [24] [27]
- Archive evidence of settings and rerun tests after the documented policy propagation window. [7] [12] [23] [24] [27]

13.5 Agents and Automation

- Use dedicated managed environments, Entra authentication, end-user credentials, channel allowlists, approved connectors, and network controls. [31] [32] [33] [34] [40] [41]
- Separate read tools from write tools. Require deterministic policy validation and human approval for material actions. [31] [32] [33] [34] [40] [41]
- For unattended work, use dedicated least-privilege service identities, limits, reconciliation, alerts, kill switches, and rollback. [31] [32] [33] [34] [40] [41]
- Block computer use in critical systems unless isolated, specially approved, and continuously monitored. [31] [32] [33] [34] [40] [41]

13.6 Logging, Retention, and Privacy

- Define retention separately for general Copilot interactions, Copilot Studio transcripts, historical activity, computer-use logs, and downstream records. [9] [10] [11] [36] [37] [38] [44]
- Scope audit, eDiscovery, DSPM, and Communication Compliance roles narrowly. Log and review searches, exports, and policy changes. [9] [10] [11] [36] [37] [38] [44]
- Document employee monitoring and privacy notices, legal holds, data-subject or customer requests, investigations, and secure deletion. [9] [10] [11] [36] [37] [38] [44]

13.7 Human Review and Use Controls

- Require source review for material factual claims. Use primary sources for legal, regulatory, policy, and customer-impacting work. [1] [51] [52] [55]
- Prohibit Copilot from being the final decision maker for credit, account restriction, SAR, employment, legal, accounting, security, or customer redress. [1] [51] [52] [55]
- Use approved templates and factual systems of record for customer communications, notices, and adverse-action reasons. [1] [51] [52] [55]
- Train users on data classes, product boundaries, hallucination, prompt injection, reporting errors, and preserving evidence. [1] [51] [52] [55]

13.8 Third-Party, Contract, and Resilience

- Map Microsoft, model providers, agents, connectors, publishers, and integrators into the third-party inventory and service chain. [23] [24] [26] [46] [47] [48]
- Review DPA, Product Terms, incident notice, audit rights, support access, subprocessors, location, deletion, business continuity, financial condition, and termination. [23] [24] [26] [46] [47] [48]
- Subscribe to product, security, subprocessor, and terms changes. Establish manual fallback for critical workflows and test provider outage and data-exit scenarios. [23] [24] [26] [46] [47] [48]

13.9 Minimum Evidence Package

Evidence	Minimum content
Risk assessment	Product, use case, data, identities, model providers, legal role, threats, controls, test results, residual risk, approvals, review date.

Evidence	Minimum content
Architecture and data flow	Prompt inputs, grounding, models, web, sources, connectors, actions, storage, logs, regions, subprocessors, outputs, deletion.
Configuration baseline	Dated tenant settings, licenses, groups, DLP, labels, web search, providers, environments, agent auth, channels, credentials, retention.
Permission evidence	Source permissions, broad-access reports, owner approvals, positive and negative persona tests, connector ACL mapping, revocation tests.
Control tests	Prompt DLP, file upload, web query, labels, eDiscovery, retention, agent actions, prompt injection, outage, kill switch, and rollback.
Vendor file	Contract, DPA, Product Terms, SOC and other reports, bridge letters, subprocessor list, incident history, financial review, BCP, exit plan.
Monitoring	Use metrics, blocked events, sensitive interactions, errors, overrides, customer-impacting output sampling, incidents, exceptions, provider changes.

14. Validation, Testing, and Continuous Monitoring

Testing should use synthetic records and canary content, not live customer data. Each test should identify the tenant setting, user persona, device, product surface, model provider, web state, source permissions, expected result, observed result, audit evidence, and remediation owner.

14.1 Pre-Production Test Domains

Domain	Pass condition
Identity and surface	Work account accepted; personal account blocked or clearly out of policy; enterprise protection visible; unlicensed behavior understood.
Permissions	Authorized persona retrieves canary; unauthorized persona cannot; external, guest, terminated, and recently removed users fail as expected.
SharePoint and connectors	Broad links, nested groups, connector ACLs, deleted records, permission changes, and revocation latency behave as designed.
Labels and encryption	Selected labels block or allow processing correctly; encrypted rights are enforced; generated response receives expected protection.
DLP typed prompt	Synthetic account, SSN, card, employee, and incident patterns trigger expected block or warning; false positives are assessed.
Direct file upload	An unlabeled synthetic sensitive file demonstrates the documented inspection gap; compensating controls stop or detect it.
Web search	Sensitive prompt does not generate web search; ordinary public research does; audit and UI behavior are captured.
Model provider	Disallowed users cannot invoke provider; allowed user can; provider change is logged; disable control takes effect.

Domain	Pass condition
Retention and eDiscovery	Prompt and response are discoverable by authorized role, held when required, and deleted according to tested policy.
Agent authentication	Unauthenticated and unapproved-channel access fails; correct user and group restrictions apply.
Agent credentials	End-user connector permissions are enforced; maker credential path is blocked or constrained; service identity scopes are minimal.
Prompt injection	Malicious instructions in documents, email, sites, and connector data cannot cause disclosure or unauthorized action.
Action safety	Approval, value limits, duplicate prevention, validation, rollback, reconciliation, and kill switch work under normal and error conditions.
Accuracy and conduct	Material samples are checked against authoritative sources, including protected-class and customer-impact scenarios.
Incident response	The bank can identify affected users, prompts, data, providers, agents, and outputs and can meet decision and notification timelines.

14.2 Monitoring Indicators

- Active users, use cases, agents, connectors, provider paths, and external channels compared with the approved inventory.
- Sensitive prompt events, web-search blocks, DLP matches, label blocks, direct uploads, and repeated policy violations.
- Broad-permission sites, anonymous or company-wide links, external users, connector ACL failures, and permission-remediation aging.
- Agent action volume, failures, retries, overrides, duplicate attempts, unusual hours, service-identity use, and kill-switch events.
- High-impact output errors, unsupported citations, customer complaints, adverse-action issues, compliance exceptions, and material rework.
- Administrative searches and exports of prompt content, role changes, provider enablement, agent publishing, environment policy changes, and retention changes.
- Microsoft roadmap, Message Center, Product Terms, DPA, subprocessor, service-health, security, and compliance-report changes.

14.3 Review Frequencies

Frequency	Review
Continuous or daily	Security alerts, service health, sensitive DLP events, high-risk agent actions, privileged changes, incidents.
Weekly	Agent failures, blocked prompts, direct uploads, user reports, provider and policy changes, open remediation.
Monthly	Inventory reconciliation, usage by tier, exceptions, action-agent samples, sensitive-content trends, training needs.

Frequency	Review
Quarterly	Access and group recertification, provider settings, web search, DLP and label tests, agent owners, connector permissions, vendor changes.
Annually	Enterprise risk assessment, board reporting, policy approval, independent testing, contract and audit review, resilience exercise.
Event driven	Major product release, model-provider change, incident, new data class, new channel or action, merger, regulatory change, material control failure.

15. Ninety-Day Implementation Road Map

Period	Objectives	Deliverables and exit criteria
Days 0-15: Contain and inventory	Establish ownership, product boundary, and current state.	Name executive and working team; inventory licenses, users, agents, connectors, model providers, web settings, DLP, retention, and high-risk sites. Prohibit consumer use and preview independent processors. Freeze new action agents until baseline exists.
Days 16-30: Set baseline	Implement identity, provider, data, and monitoring controls.	Work-account and device policy; provider allowlist; web-search rule; initial DLP and labels; admin-role review; SharePoint risk reports; retention decision; agent environment policy; user and help-desk training.
Days 31-60: Remediate and pilot	Use a small, representative, noncritical pilot.	Fix priority oversharing; run positive and negative permission tests; test upload gap, web, labels, provider, audit, retention, and hallucination; approve Tier 1 and selected Tier 2 use cases; collect evidence and user incidents.
Days 61-90: Govern and scale	Move from pilot to controlled production.	Risk committee approval; finalized policy and exception process; vendor due diligence; monitoring dashboard; quarterly recertification; incident tabletop; agent approval packet; board or committee reporting; go or no-go decision by data class.

Deployment gate

Do not equate license purchase or tenant availability with production approval. A user group should receive access only after the product, provider, web, permissions, DLP, retention, monitoring, and use-case evidence for that group is complete.

16. Vendor Due-Diligence and Contract Checklist

Service and Data

- Confirm: Exact Copilot products, licenses, model providers, agents, connectors, channels, regions, and features in scope. [1] [5] [8] [9] [23] [24]
- Confirm: All customer-data categories processed, including prompt text, Graph excerpts, files, metadata, responses, citations, feedback, logs, transcripts, screenshots, support records, and generated Bing queries. [1] [5] [8] [9] [23] [24]

- Confirm: Data-at-rest, transient-processing, backup, disaster-recovery, telemetry, and support locations by provider path. [1] [5] [8] [9] [23] [24]
- Confirm: Retention and deletion for active data, backups, holds, safety systems, feedback, model-provider data, Dataverse transcripts, and audit records. [1] [5] [8] [9] [23] [24]

Legal and Privacy

- Confirm: Microsoft DPA and Product Terms applicability to each feature and clear identification of any separate Bing or independent-processor terms. [2] [7] [23] [24] [26]
- Confirm: Processor instructions, purpose limitation, confidentiality, government requests, data-subject support, cross-border mechanism, and deletion assistance. [2] [7] [23] [24] [26]
- Confirm: Subprocessor list, advance-change notice, objection or termination rights, and feature behavior when a provider is disabled. [2] [7] [23] [24] [26]

Security and Assurance

- Confirm: Encryption, key management, tenant isolation, secure development, vulnerability management, penetration testing, access controls, logging, monitoring, and incident response. [2] [14] [15] [23]
- Confirm: SOC 1 and SOC 2, ISO, PCI, FedRAMP, HITRUST, bridge letters, exceptions, complementary user-entity controls, and scope by actual model-provider path. [2] [14] [15] [23]
- Confirm: Support access, Customer Lockbox, administrator access, emergency access, personnel screening, and subcontractor controls. [2] [14] [15] [23]

Operational Resilience

- Confirm: Service levels, capacity, rate limits, change windows, material degradation, regional failover, backup, restoration, dependency mapping, and testing. [47] [48] [56]
- Confirm: Incident-notice trigger, timeline, contacts, required content, ongoing updates, forensic evidence, root cause, corrective action, and bank regulatory deadlines. [47] [48] [56]
- Confirm: Business continuity, exit assistance, data export, agent and configuration portability, secure deletion certificate, and transition support. [47] [48] [56]

Governance and Change

- Confirm: Product roadmap, preview and general-availability definitions, model updates, evaluation evidence, deprecation, customer notice, and rollback options. [7] [23] [24] [27] [31] [32]
- Confirm: Admin controls for web search, model providers, agents, connectors, feedback, retention, audit, channels, credentials, and external sharing. [7] [23] [24] [27] [31] [32]
- Confirm: Right to receive relevant examination support, audit reports, regulator information requests, and material control-change evidence. [7] [23] [24] [27] [31] [32]

16.1 Questions for Microsoft or the Reseller

1. For each licensed feature, which entity operates the model and what administrative evidence identifies the provider used for a specific interaction?

2. Which features are enabled by default today, and what future enablement can occur without an explicit tenant opt-in?
3. Exactly what does Zero Data Retention cover for OpenAI-operated models, and which feature limitations, safety systems, or logs are outside it?
4. Which Copilot data and provider paths are in scope for each SOC, ISO, PCI, FedRAMP, and HITRUST report supplied to the bank?
5. What prompt and response data can Microsoft personnel, provider personnel, support engineers, or automated safety systems access, under which conditions, and with what logs and approvals?
6. How are bank instructions, deletions, legal holds, eDiscovery, backups, safety retention, and provider deletion reconciled?
7. What evidence can the bank receive after a suspected data disclosure, prompt injection, unsafe action, permission-trimming failure, or provider incident?
8. How quickly do provider, web, DLP, permission, retention, and agent-policy changes propagate, and what control is available during propagation?
9. Can the bank export a complete inventory and configuration history for Copilot users, agents, providers, connectors, knowledge sources, credentials, channels, and environments?
10. What assistance is available for regulator examinations, customer-notice analysis, SAR-related investigations, litigation holds, and secure service termination?

17. Policy Language Starter

Use of this section

The following language is a starting point for bank policy drafting. Legal, privacy, compliance, information security, records, HR, model-risk, and business owners should adapt and approve it.

Approved services

Employees may use only bank-approved Copilot services while signed in with their bank-managed work identity and, for sensitive use, from a bank-managed device. Consumer accounts, personal browser profiles, unapproved plugins, preview models, independent-processor models, agents, and external channels may not receive bank information unless specifically approved.

Data minimization

Users must provide only the minimum information necessary for the approved business purpose. Authentication secrets, passwords, access tokens, private keys, sensitive authentication data, and SAR or SAR-revealing information are prohibited. Customer NPI, employee restricted data, legal, fraud, security, and payment data require the approval defined by the data-class matrix.

Web grounding

Public web search may be used for public and approved internal research. Users must not enable or invoke web search with restricted data. The bank may block web search based on user, group, agent, prompt content, or data classification.

Accuracy and decisions

Copilot output is a draft or analytical aid unless an approved use-case standard states otherwise. Users must verify material facts against authoritative sources. Copilot may not make final credit, BSA/AML, account, legal, employment, accounting, security, or customer-redress decisions.

Customer communications

Generated customer communications must use approved facts, disclosures, templates, and review. Copilot may not invent or substitute adverse-action reasons, contractual terms, fees, rates, rights, deadlines, or regulatory statements.

Agents and actions

Agents must be inventoried, risk-rated, tested, and approved before publication. Entra authentication and end-user credentials are the default. Maker credentials, external channels, event triggers, autonomous actions, and computer use require enhanced approval and technical safeguards.

Records and monitoring

Copilot interactions may be logged, retained, reviewed, and disclosed for security, compliance, legal, records, quality, and regulatory purposes. Users should not enter information inconsistent with approved purpose or need-to-know restrictions. Administrative access to interaction content is limited and monitored.

Reporting

Users must promptly report suspected disclosure, inaccurate material output, unsafe action, permission error, prompt injection, consumer-account use, or unapproved provider use. The bank may suspend access while an event is investigated.

18. Final Risk Position

Microsoft provides a credible enterprise protection framework for Copilot: contractual processing commitments, no foundation-model training on organizational prompts and responses, encryption, tenant isolation, permission-aware grounding, Purview controls, retention and eDiscovery integration, and administrative settings. Those protections support bank adoption. [1] [2] [3] [4] [9] [10]

The remaining risk is material and configuration-dependent. Copilot can make overshared data easier to find, store sensitive interaction records, create a separate Bing query flow, use model providers with different geography and certifications, and power agents that act with credentials and external

channels. DLP has important gaps, particularly for directly uploaded files. A bank cannot approve 'Copilot' generically and expect the brand to define the boundary. [5] [7] [12] [17] [22] [23] [24] [33] [40]

The recommended position is conditional approval for enterprise work-account use, phased by data class and use-case tier. Consumer Copilot is prohibited for bank data. Copilot Studio begins as a controlled read-only pilot. Action, autonomous, scheduled, independent-processor, and computer-use configurations require enhanced review. SAR content, secrets, full payment-card data, and unreviewed material decisions remain prohibited absent a specialized and separately approved architecture. [31] [40] [50] [51] [52] [58]

Go decision

Proceed only when the bank can answer, with evidence: who can use the service, which data it can reach, which provider processes it, whether web search is active, what is stored, who can review it, how long it remains, what an agent can do, how failures are detected, and how access or automation can be stopped.

Appendix A. Source Notes

Sources were accessed August 20, 2026. Product features, names, rollout state, provider enablement, contract terms, and regulatory guidance can change. Use the links below to validate the bank's current tenant and legal position. Microsoft sources describe product behavior and commitments; regulator and standards sources support the bank-risk mapping.

[1] Data, Privacy, and Security for Microsoft Copilot. Microsoft Learn. Accessed August 20, 2026.

Core Microsoft statement on prompts, responses, Microsoft Graph grounding, training, storage, model processing, feedback, agents, and web search.

[2] Enterprise Data Protection for Microsoft Copilot and Copilot Chat. Microsoft Learn. Accessed August 20, 2026.

Explains the Microsoft Products and Services Data Protection Addendum, Product Terms, encryption, tenant isolation, and the separate Bing web-search boundary.

[3] Microsoft Copilot Architecture. Microsoft Learn. Accessed August 20, 2026.

Describes orchestration, grounding with Microsoft Graph, permissions, encrypted data flow, and interaction history.

[4] Data Protection and Auditing Architecture for Microsoft Copilot. Microsoft Learn. Accessed August 20, 2026.

Describes sensitivity labels, encryption usage rights, SharePoint controls, auditing, and interaction storage.

[5] Privacy and Protections in Microsoft Copilot Chat. Microsoft Learn. Accessed August 20, 2026.

Details Copilot Chat grounding, Exchange-based interaction storage, Bing query handling, regional processing, and organizational data access paths.

[6] Microsoft Copilot Chat Overview. Microsoft Learn. Accessed August 20, 2026.

Compares Copilot Chat and licensed Microsoft Copilot capabilities, including web grounding, file upload, agents, and work-data grounding.

[7] Manage Web Search in Microsoft Copilot. Microsoft Learn. Accessed August 20, 2026.

Explains generated Bing search queries, identifier removal, admin controls, and limitations of the web-search data flow.

[8] Data Residency for Microsoft Copilot. Microsoft Learn. Accessed August 20, 2026.

Describes data-at-rest location for interaction content and semantic index information under Microsoft 365 residency commitments.

[9] Retention Policies for Microsoft Copilot. Microsoft Learn. Accessed August 20, 2026.

Explains Exchange mailbox storage, hidden folders, retention timing, holds, and disposition of prompts and responses.

[10] Microsoft Purview Data Security and Compliance Protections for Copilot. Microsoft Learn. Accessed August 20, 2026.

Describes auditing, eDiscovery, DSPM for AI, sensitivity labels, classifiers, and communication compliance.

[11] Search for and Delete Microsoft Copilot Data in eDiscovery. Microsoft Learn. Accessed August 20, 2026.

Explains how compliance teams locate, preserve, review, and purge Copilot interaction data.

[12] Data Loss Prevention for Microsoft Copilot. Microsoft Learn. Accessed August 20, 2026.

Documents DLP controls and important coverage limits, including direct file-upload inspection limits, supported content, and policy propagation time.

[13] Learn About Sensitivity Labels. Microsoft Learn. Accessed August 20, 2026.

Explains labels, encryption, protection persistence, and Copilot recognition of labeled content.

[14] Customer Key Overview. Microsoft Learn. Accessed August 20, 2026.

Documents customer-controlled encryption-key coverage, including Microsoft Copilot interactions at rest.

[15] Customer Lockbox Requests. Microsoft Learn. Accessed August 20, 2026.

Describes customer approval controls for Microsoft support access to covered Microsoft 365 content, including Copilot interactions through Exchange Online.

[16] Restricted Content Discovery. Microsoft Learn. Accessed August 20, 2026.

Describes a temporary control that reduces organization-wide discovery of selected SharePoint content without changing permissions.

[17] Restricted SharePoint Search. Microsoft Learn. Accessed August 20, 2026.

Explains a temporary Copilot and search restriction, including its site limit and the fact that it is not a security boundary.

[18] SharePoint Advanced Management. Microsoft Learn. Accessed August 20, 2026.

Overview of SharePoint governance controls and reports that can reduce oversharing before broad Copilot deployment.

[19] Site Access Review. Microsoft Learn. Accessed August 20, 2026.

Describes owner-led review of site access and oversharing findings.

[20] Data Access Governance Reports. Microsoft Learn. Accessed August 20, 2026.

Describes reports for broad permissions, sharing links, and other access risks.

[21] Restricted Access Control for SharePoint and OneDrive. Microsoft Learn. Accessed August 20, 2026.

Describes an access boundary that can require both content permission and membership in an allowed group.

[22] Manage Access Permissions for Microsoft Copilot Connectors. Microsoft Learn. Accessed August 20, 2026.

Warns that connector permission mappings must match source permissions and that broad mappings can expose indexed content.

[23] OpenAI as a Subprocessor for Microsoft Copilot. Microsoft Learn. Accessed August 20, 2026.

Current 2026 terms, enablement, controls, Zero Data Retention configuration, geography exclusions, and compliance-attestation exclusions for OpenAI-operated models.

[24] Connect to Anthropic AI Models as a Subprocessor. Microsoft Learn. Accessed August 20, 2026.

Explains Anthropic subprocessor and independent-processor paths, regional exclusions, admin controls, and data retention for selected preview models.

[25] AI Models in Microsoft Copilot. Microsoft Learn. Accessed August 20, 2026.

Defines Microsoft-hosted models, subprocessors under Microsoft terms, and independent processors under separate terms.

[26] Microsoft Copilot Subprocessor Overview. Microsoft Learn. Accessed August 20, 2026.

Explains supplier assessment, contractual restrictions, ongoing monitoring, and the Trust Center subprocessor list.

[27] Manage Preview AI Models. Microsoft Learn. Accessed August 20, 2026.

Explains risks and administrative controls for preview models, including models that retain data outside Microsoft geography commitments.

[28] Microsoft Copilot Application Card. Microsoft Learn. Accessed August 20, 2026.

Central product card covering model providers, security, privacy, compliance, and service dependencies.

[29] Privacy FAQ for Consumer Microsoft Copilot. Microsoft Support. Accessed August 20, 2026.

Describes consumer Copilot privacy choices and model-training controls. It is not the enterprise-data-protection commitment.

[30] Conversation History in Consumer Microsoft Copilot. Microsoft Support. Accessed August 20, 2026.

Describes consumer conversation-history behavior and retention.

[31] Security and Governance in Microsoft Copilot Studio. Microsoft Learn. Accessed August 20, 2026.

Copilot Studio identity, environments, governance, channels, data policies, and security overview.

[32] Use Data Policies to Govern Copilot Studio. Microsoft Learn. Accessed August 20, 2026.

Describes Power Platform data policies that can block connectors, knowledge sources, channels, HTTP, triggers, and unauthenticated publishing.

[33] Configure Agent Authentication and Block No Authentication. Microsoft Learn. Accessed August 20, 2026.

Documents maker and end-user credentials, no-authentication controls, and the risk of author credentials granting unintended reach.

[34] Configure End-User Authentication. Microsoft Learn. Accessed August 20, 2026.

Explains connection authorization, token handling, revocation, and identity-aware access for agent actions.

[35] Add Knowledge to a Copilot Studio Agent. Microsoft Learn. Accessed August 20, 2026.

Explains knowledge sources, user permission trimming for selected sources, uploaded documents in Dataverse, and ungrounded response settings.

[36] Manage Agent Activity Data in Microsoft 365. Microsoft Learn. Accessed August 20, 2026.

Describes storage and compliance management for Copilot Studio agent activity in Microsoft 365 services.

[37] Custom Analytics Strategy for Copilot Studio. Microsoft Learn. Accessed August 20, 2026.

Documents Dataverse conversation transcripts and their default retention behavior.

[38] Audit Logging for Copilot Studio. Microsoft Learn. Accessed August 20, 2026.

Explains audit event metadata, interaction content in Purview, and retention considerations.

[39] Review Agent Activity. Microsoft Learn. Accessed August 20, 2026.

Shows that activity review can include user input, trigger payloads, responses, and reasoning information for supported models.

[40] Computer Use in Copilot Studio. Microsoft Learn. Accessed August 20, 2026.

Describes computer-use agents, credential behavior, machine access, and controls for automated user-interface actions.

[41] Security and Governance Phase 3 for Copilot Studio. Microsoft Learn. Accessed August 20, 2026.

Detailed deployment controls for environments, groups, channels, connectors, network isolation, and monitoring.

[42] Data Privacy and Security for Web Search in Copilot Studio. Microsoft Learn. Accessed August 20, 2026.

Explains generated Bing queries and the separate legal terms that apply to Bing web-search data.

[43] Use Microsoft Purview with Copilot Studio. Microsoft Learn. Accessed August 20, 2026.

Explains Purview coverage and channel or licensing caveats for Copilot Studio interactions.

[44] Monitor Computer Use. Microsoft Learn. Accessed August 20, 2026.

Documents transcript, screenshot, and diagnostic-log collection for computer-use activity.

[45] Zero Trust for Microsoft Copilot. Microsoft Learn. Accessed August 20, 2026.

Maps identity, devices, applications, data, collaboration, threat protection, and permissions into a Copilot security approach.

[46] Appendix B to 12 CFR Part 364: Interagency Guidelines Establishing Information Security Standards. Electronic Code of Federal Regulations. Accessed August 20, 2026.

Requires a written information-security program, risk assessment, safeguards, testing, service-provider oversight, incident response, and board reporting for FDIC-supervised institutions.

[47] Interagency Guidance on Third-Party Relationships: Risk Management. Office of the Comptroller of the Currency. Accessed August 20, 2026.

Applies a risk-based lifecycle of planning, due diligence, contract negotiation, monitoring, and termination to bank third-party relationships.

[48] FFIEC Statement on Security in a Cloud Computing Environment. Federal Financial Institutions Examination Council. Accessed August 20, 2026.

Emphasizes shared responsibility, due diligence, configuration, monitoring, and protection of sensitive customer information in cloud services.

[49] Regulation P, Privacy of Consumer Financial Information. Electronic Code of Federal Regulations. Accessed August 20, 2026.

Governs notices and limits on disclosure of nonpublic personal information, including service-provider and transaction-processing exceptions.

[50] 31 CFR 1020.320, Reports by Banks of Suspicious Transactions. Electronic Code of Federal Regulations. Accessed August 20, 2026.

Requires SAR record retention and strictly limits disclosure of a SAR or information revealing its existence.

[51] CFPB Circular 2022-03: Adverse Action and Complex Algorithms. Consumer Financial Protection Bureau. Accessed August 20, 2026.

States that ECOA and Regulation B adverse-action requirements apply regardless of algorithmic complexity.

[52] CFPB Circular 2023-03: Specific Reasons for Adverse Action. Consumer Financial Protection Bureau. Accessed August 20, 2026.

Emphasizes accurate, specific principal reasons when AI or other predictive methods influence credit decisions.

[53] SR 26-2: Revised Guidance on Model Risk Management. Board of Governors of the Federal Reserve System. Accessed August 20, 2026.

Current 2026 interagency model-risk guidance. It expressly excludes generative and agentic AI from scope while directing banks to use risk-management and governance practices for tools outside scope.

[54] Artificial Intelligence Risk Management Framework 1.0. National Institute of Standards and Technology. Accessed August 20, 2026.

Voluntary AI governance framework organized around Govern, Map, Measure, and Manage.

[55] NIST AI 600-1: Generative Artificial Intelligence Profile. National Institute of Standards and Technology. Accessed August 20, 2026.

Adds generative-AI risks and actions covering confabulation, privacy, information integrity, human oversight, incident disclosure, and third-party risk.

[56] 12 CFR Part 53: Computer-Security Incident Notification. Electronic Code of Federal Regulations. Accessed August 20, 2026.

For OCC-supervised banks, requires qualifying incident notification within 36 hours and specified service-provider notice for material disruption. Parallel rules apply to Board and FDIC institutions.

[57] Bank Service Company Act Notice. Federal Deposit Insurance Corporation. Accessed August 20, 2026.

Explains written notice obligations for covered service relationships under the Bank Service Company Act.

[58] PCI Data Security Standard. PCI Security Standards Council. Accessed August 20, 2026.

Provides baseline technical and operational requirements for environments that store, process, or transmit payment account data.

[59] AI and Payments: Exploring Pitfalls and Potential Security Risks. PCI Security Standards Council. Accessed August 20, 2026.

Highlights questions about sensitive data, credentials, AI-provider handling, security scope, and human review in payment environments.

Appendix B. Detailed Test Scripts

B.1 Permission trimming

1. Create synthetic files in a restricted site, a broad site, a user's OneDrive, and a connector source. Include unique canary phrases and no real customer data. [3] [17] [20] [21] [22]
2. Assign personas: authorized employee, unauthorized employee, external guest, recently removed user, terminated user, and administrator without business access. [3] [17] [20] [21] [22]
3. Ask equivalent questions in licensed Copilot, Copilot Chat, and each approved agent. Record citations, excerpts, response labels, and audit events. [3] [17] [20] [21] [22]
4. Change source permissions and group membership. Measure revocation latency and confirm previous chat context does not continue to disclose the canary. [3] [17] [20] [21] [22]
5. Pass only if every unauthorized persona fails, every authorized persona receives only expected content, and evidence explains any cache or propagation delay. [3] [17] [20] [21] [22]

B.2 DLP and direct upload

1. Create synthetic prompts containing test SSNs, account numbers, card patterns, employee identifiers, incident terms, and restricted project names. [12] [13]
2. Verify block, warning, web-search suppression, audit, user message, and policy-match evidence for typed prompts. [12] [13]
3. Create labeled and unlabeled files containing the same synthetic patterns. Test upload, open-file context, SharePoint grounding, and connector retrieval separately. [12] [13]
4. Document the direct-upload content-inspection gap and prove that mandatory labeling, endpoint, browser, repository, or monitoring controls compensate. [12] [13]
5. Repeat after the maximum documented policy propagation period and after any licensing or DLP rule change. [12] [13]

B.3 Web search

1. Use a public research prompt and confirm a generated web search occurs with appropriate citations. [5] [7] [12] [42]
2. Use a synthetic sensitive prompt with web search allowed, then inspect available audit and UI evidence to understand the query path. [5] [7] [12] [42]
3. Apply the bank's web-search restriction and repeat. Confirm no Bing query is issued for the protected scenario. [5] [7] [12] [42]
4. Test sensitive terms supplied through typed text, an uploaded file, Outlook context, and an agent knowledge source because input paths can differ. [5] [7] [12] [42]
5. Pass only if the bank can explain which data enters Bing, under which terms, and how the block is evidenced. [5] [7] [12] [42]

B.4 Provider control

1. Export the current model-provider and preview-model settings, eligible groups, and defaults. [23] [24] [25] [26] [27]

2. Use an allowed test user and a disallowed user. Confirm feature availability and provider path match policy. [23] [24] [25] [26] [27]
3. Disable the provider. Measure propagation and verify that existing chats, agents, and features no longer invoke it. [23] [24] [25] [26] [27]
4. Compare data class, region, retention, contract, and assurance reports for Microsoft-hosted, subprocessor, and independent-processor paths. [23] [24] [25] [26] [27]
5. Create an alert and change ticket process for any default, provider, or terms change. [23] [24] [25] [26] [27]

B.5 Agent identity and credentials

1. Publish only in a test environment. Attempt access as an authenticated allowed user, authenticated disallowed user, external user, and anonymous user. [31] [32] [33] [34]
2. Connect a synthetic source with different records for different personas. Verify end-user permission trimming. [31] [32] [33] [34]
3. Switch to maker credentials only in a controlled test and prove the privilege increase. Then block or constrain that mode according to policy. [31] [32] [33] [34]
4. Test service-identity token expiry, revocation, group removal, owner departure, secret rotation, and emergency disable. [31] [32] [33] [34]
5. Pass only if identity and action authority are deterministic, documented, auditable, and recoverable. [31] [32] [33] [34]

B.6 Prompt injection and actions

1. Place malicious instructions in a synthetic email, document, website, support ticket, and connector record. Examples should ask the agent to ignore policy, reveal another record, or execute an unrelated action. [40] [41] [55]
2. Test read-only and action-enabled versions. Confirm retrieved instructions are treated as untrusted content. [40] [41] [55]
3. Attempt parameter manipulation, hidden instructions, long-context distraction, repeated retries, duplicate submission, and approval bypass. [40] [41] [55]
4. Verify independent validation, allowlists, confirmation, limits, idempotency, reconciliation, rollback, kill switch, and alerts. [40] [41] [55]
5. Record every successful or partially successful attack as a control failure until remediated and retested. [40] [41] [55]

B.7 Retention, audit, and deletion

1. Create uniquely tagged interactions in each approved surface and agent. Include no real customer data. [9] [10] [11] [36] [37] [38] [44]
2. Confirm user history, Exchange or Microsoft 365 compliance location, Dataverse transcript, audit event, DSPM view, and any computer-use log. [9] [10] [11] [36] [37] [38] [44]
3. Apply retention, legal hold, user deletion, eDiscovery search, export, and approved purge steps. Measure timing and document every surviving copy. [9] [10] [11] [36] [37] [38] [44]

4. Verify least-privilege roles and that searches and exports themselves are logged and reviewed. [9] [10] [11] [36] [37] [38] [44]
5. Pass only when legal, records, privacy, security, and audit agree that the lifecycle matches policy. [9] [10] [11] [36] [37] [38] [44]

Appendix C. Control Ownership RACI

R = responsible, A = accountable, C = consulted, I = informed. Titles should be mapped to the bank's actual organization. No single function can own this risk alone.

Control area	Business	IT / M365	Security	Privacy / Legal / Compliance	Records	Audit
Use-case purpose and outcome	A/R	C	C	C	I	I
Tenant, identity, license, and provider settings	I	A/R	C	C	I	I
Data classification and lawful use	R	C	C	A/R	C	I
SharePoint and connector permissions	A/R	R	C	C	I	I
DLP, labels, web, and monitoring	C	R	A/R	C	C	I
Agent design and action safety	A/R	R	C	C	I	I
Retention, hold, and deletion	C	R	C	C	A/R	I
Third-party due diligence and contract	C	C	C	A/R	I	I
Testing and risk acceptance	A	R	R	R	C	C
Independent assurance	I	C	C	C	C	A/R
Incident response and notification	C	R	A/R	R	C	I
Board and committee reporting	A/R	C	C	C	C	I

C.1 Required Decision Owners

- The data owner approves which data classes and repositories may ground the use case.
- The business owner accepts the outcome and customer or operational risk and confirms the human-review design.
- Technology owns tenant and agent configuration, availability, change control, and lifecycle operations.
- Security owns threat modeling, identity, DLP, monitoring, incident response, and technical testing standards.
- Privacy, legal, and compliance determine legal basis, notices, restrictions, regulatory mapping, and exception conditions.
- Records owns retention, hold, disposition, and the relationship between Copilot interactions and downstream records.
- Internal audit independently assesses whether governance and controls are designed and operating effectively.

End of report