



AI Search SharePoint Setup: Pre-Flight Readiness Checklist

Use this audit framework to validate your SharePoint environment before connecting Azure AI Search. A clean architecture guarantees high-trust generative AI answers.

Phase 1: Content Hygiene (The "Junk" Audit)

- **Identify Authoritative Sources:** Selected a single, highly curated Document Library for the initial pilot (avoid indexing the entire tenant).
- **Purge ROT (Redundant, Outdated, Trivial):** Archived or deleted duplicate files, draft versions, and legacy documents older than the retention policy allows.
- **Verify File Formats:** Confirmed that the target library contains supported file types (e.g., standard Office documents, readable PDFs) and flagged unsupported or password-protected files.

Phase 2: Metadata & Taxonomy

- **Standardize Columns:** Replaced generic folder structures with strict metadata columns (e.g., *Document Type*, *Department*, *Effective Date*).
- **Enforce Naming Conventions:** Ensured file names are descriptive and human-readable (e.g., *Q3_2026_Marketing_Budget.xlsx* instead of *Book1_final_v2.xlsx*).
- **Map Retrievable Fields:** Documented exactly which SharePoint columns provide semantic value and should be mapped to the Azure AI Search index schema.

Phase 3: Permissions & Security (RBAC & ACL)

- **Audit Site Ownership:** Verified active owners for the target SharePoint site and removed orphaned user accounts.
- **Validate Access Control Lists (ACLs):** Confirmed that document-level and library-level permissions are strict and accurate (no accidental "Everyone except external users" access on sensitive files).
- **Define App Identity:** Created and securely stored the App Registration, Application ID, and Tenant ID required for the Azure AI Search SharePoint indexer connection string.



Phase 4: Indexing Strategy

- **Define the Scope:** Clearly documented the use case (e.g., "HR Policy Q&A Retrieval" vs. "Enterprise Portal Search").
- **Plan Deletion Handling:** Verified that the indexer is configured to automatically detect and drop deleted SharePoint files from the search index.
- **Establish a Refresh Schedule:** Determined the frequency of incremental indexing based on how often the source library is updated.