# Lab 3 – Knowledge Indexing & Retrieval (Unstructured PDFs)

This lab focuses on transforming unstructured underwriting and claims manuals into AI-searchable knowledge. Learners will build an end-to-end retrieval pipeline using underwriting and claims PDF documents, enabling downstream agentic reasoning and decision support.

**Lab Goals**

• Understand the role of unstructured documents (underwriting manuals, claims manuals, NFIP guidance) in underwriting AI systems
• Ingest provided underwriting and claims PDF manuals into Microsoft Fabric / OneLake
• Design an unstructured document processing pipeline (PDF to text to chunks to embeddings)
• Create an Azure AI Search index for unstructured underwriting knowledge
• Generate vector embeddings for document chunks using Azure OpenAI
• Configure hybrid retrieval combining semantic search, vector search, and metadata filtering
• Validate retrieval quality using underwriting and compliance questions

**Hands-On Activities**

- Upload underwriting and claims PDF manuals into Fabric OneLake storage
- Create and populate an Azure AI Search index with PDF manuals. Apply document chunking and generate vector embeddings.
- Test search and retrieval results using sample underwriting queries

**Dependencies and Prerequisites**

- Microsoft Fabric workspace (Contributor access)
- OneLake storage enabled
- Azure AI Search service
- Azure OpenAI resource (embeddings model access)
- Provided underwriting and claims PDF manuals

**Outputs of This Lab**

- AI-searchable index containing underwriting and claims knowledge
- AI-ready data foundation for search, retrieval, and agentic reasoning

# Hands-On Activities: Step by step instructions

## 1. Download structured and unstructured data for Underwriting solution

[Skip step 1 if you have completed download step] Download following datasets onto your laptop from the provided links. We will be uploading them to OneLake in the next steps.

| Name / Source | Description | Link |
|---|---|---|
| **FIMA NFIP Redacted Claims v2 (FEMA)** | Over 2.7 million flood-insurance claim transactions. Ideal for modelling peril-specific property risk and exposure. | FIMA NFIP Redacted Claims - v2 \| FEMA.gov |
| **Texas FAIR Plan Underwriting Manual** | Official underwriting manual for residential property coverage under the Texas FAIR Plan — details eligibility, coverage, inspection, and risk rules. | TFPA-Underwriting-Manual_Edition-Date-04-2023.pdf |
| **NFIP Claims Manual** | FEMA's official claims-handling manual under the National Flood Insurance Program — valuable for flood-risk decision logic and claims-process transparency. | NFIP Claims Manual (June 2025) |

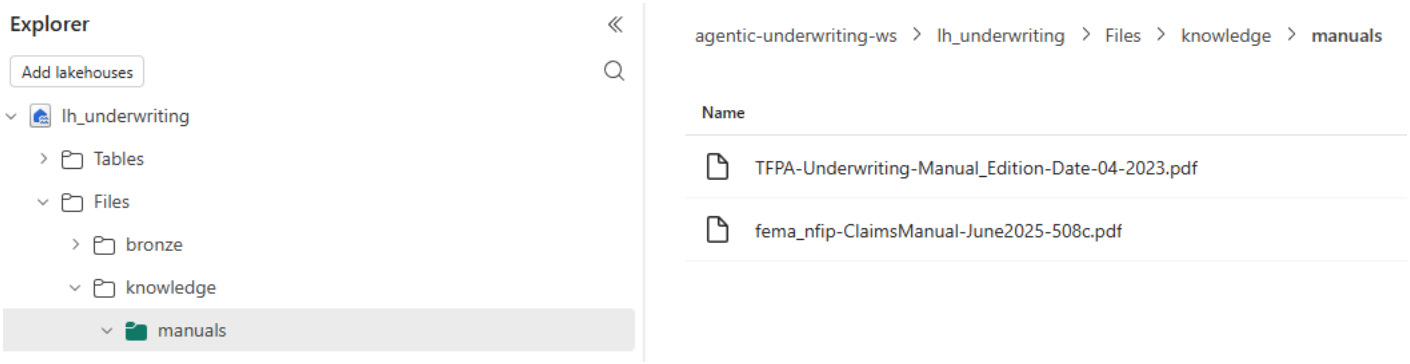
## 2. Upload PDF Data to Fabric OneLake

Purpose: Store the unstructured data in Fabric.

CAUTION: Ensure PDF manuals does not have purview sensitivity labels before you upload them to Fabric OneLake. This will cause AI search indexer to fail so remove sensitivity labels, and then upload to Fabric.

Instructions:

1. Login to your Fabric workspace
2. Create a "knowledge" folder under Files.
3. Upload PDF dataset under "manuals" folder.
4. Verify upload completion. It should look as below:




## 3. Connect Azure AI Search to Fabric OneLake

Purpose: Enable Azure AI Search service access to the unstructured data stored in Fabric OneLake.

To allow a Managed Identity to access Fabric OneLake, you must convert it into an Entra Enterprise Application (which Fabric *can* assign permissions to). Here is how to add Azure AI Search's System Assigned Managed Identity into Fabric

**STEP 1 — Get the Azure AI Search Managed Identity Object ID**

Azure Portal → *Azure AI Search* → **Identity**
Copy the **Object (principal) ID**.

**STEP 2 — Find the Managed Identity in Entra ID**

Even though it doesn't appear in the Fabric UI picker, *it DOES exist* in Entra ID as an Enterprise App.

1. Go to **Microsoft Entra ID → Enterprise Applications**
2. Click **All Applications**
3. In the search box, paste the **Object ID** (or search by name: AzureSearch-<yoursearchservice>)
4. You should now see an app such as:

   AzureSearch-yoursearchservicename

   Type: Managed Identity

**STEP 3 — Assign a Role to this Managed Identity in Fabric**

Now that you can see the MI as an Enterprise App, Fabric will allow assigning it.

1. Go to **Fabric → Workspace**
2. Open **Manage Access**
3. Click **Add people or groups**

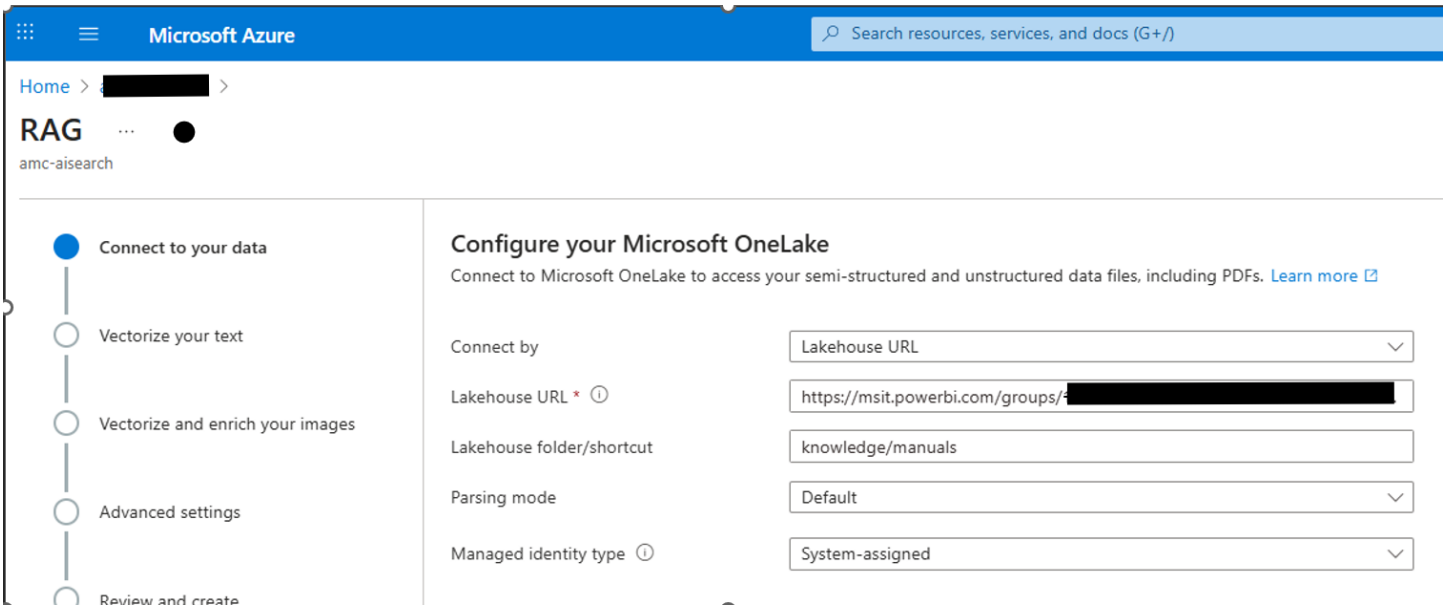
4. In the search box, paste the MI's **display name** you found under Enterprise Apps
   Example:
5. AzureSearch-yoursearchservice
6. Select it
7. Assign role: **Contributor** (required for listing items in a Lakehouse)

**STEP 4 — Grant permissions directly on the Lakehouse (required)**

1. Open the **Lakehouse**
2. Click **More (...) → Permissions**
3. Add the SAME identity
4. Assign at minimum:
   - **Read All**
   - **Read SQL endpoint**

**STEP 5 — Connect Azure AI Search to Microsoft OneLake to index underwriting manuals (pdf)**

- Go to Azure Portal ➔ Azure AI Search
- Go to Add Data option.
- Note: You might need to create a stand-alone Azure OpenAI service instance if your Foundry AOAI doesn't support system assigned ID for authentication. Give Azure AI Search Managed Identity access to the Azure OpenAI resource.
- Configure Connect to your data using your OneLake details.



With that, you should have AI Search with documents indexed directly from OneLake.