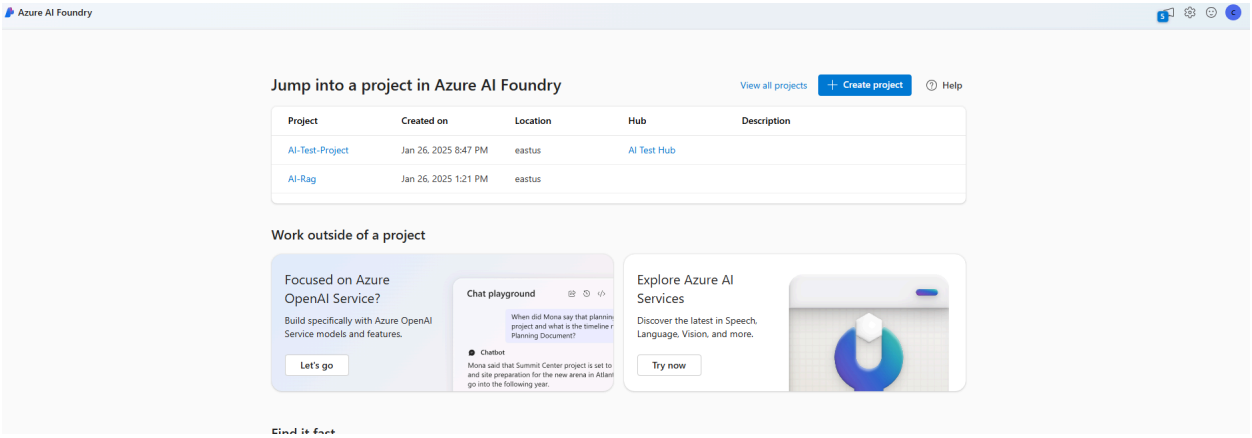
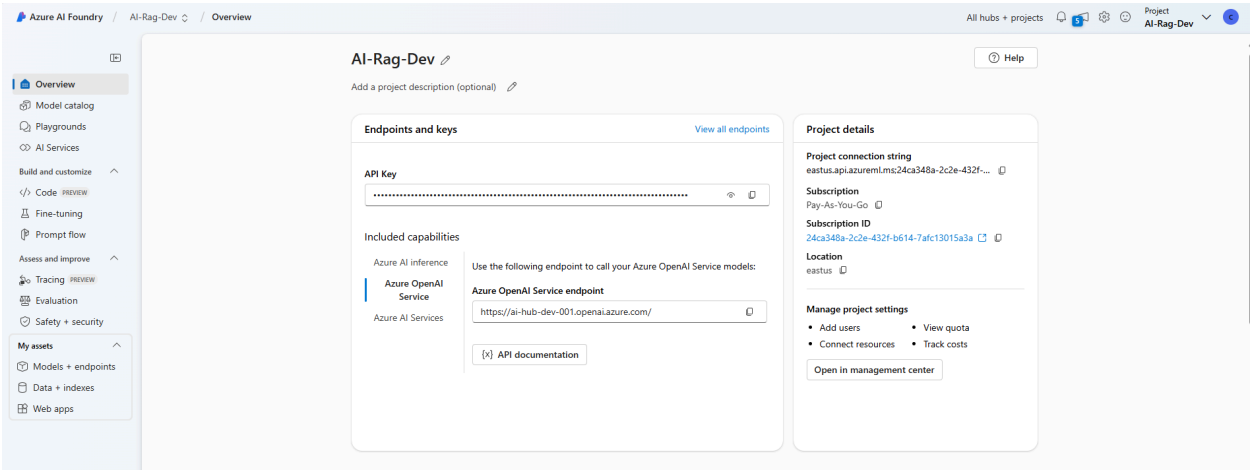
Set up resource group for Hub

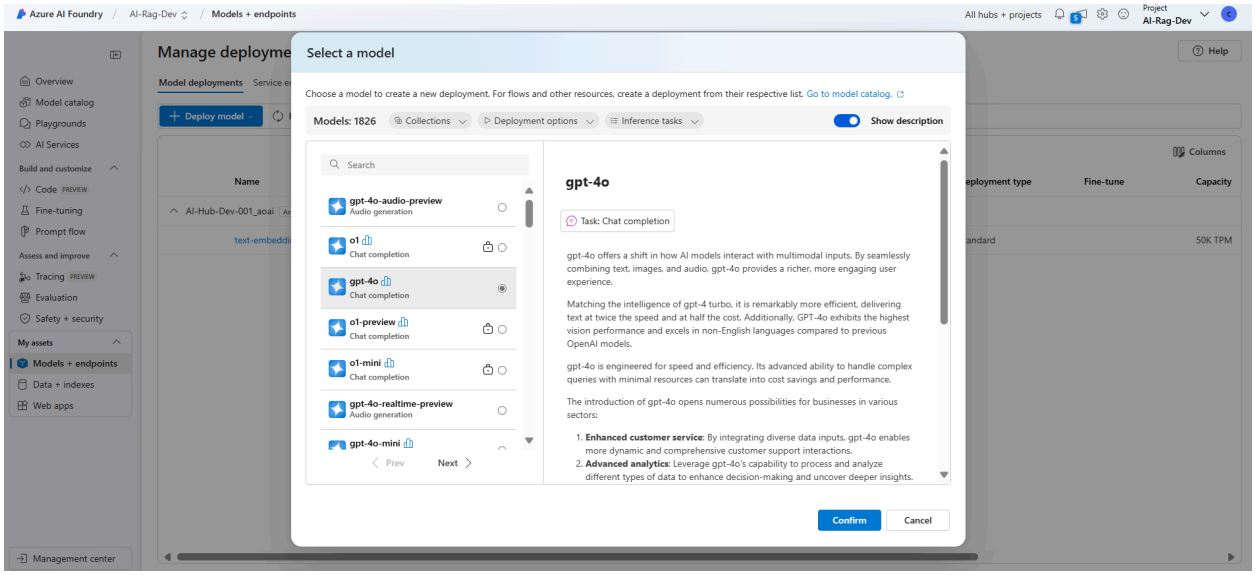
Create AI hub under Azure AI Foundry. This will include storage, networking, encryption, identity and tags. Note the networking for all projects under the hub will share the same networking. If you get a networking error during creation you will need to have a unique name for the “Connect AI Services incl. OpenAI” property. Once resource is create navigate to https://ai.azure.com/ (hub dashboard)



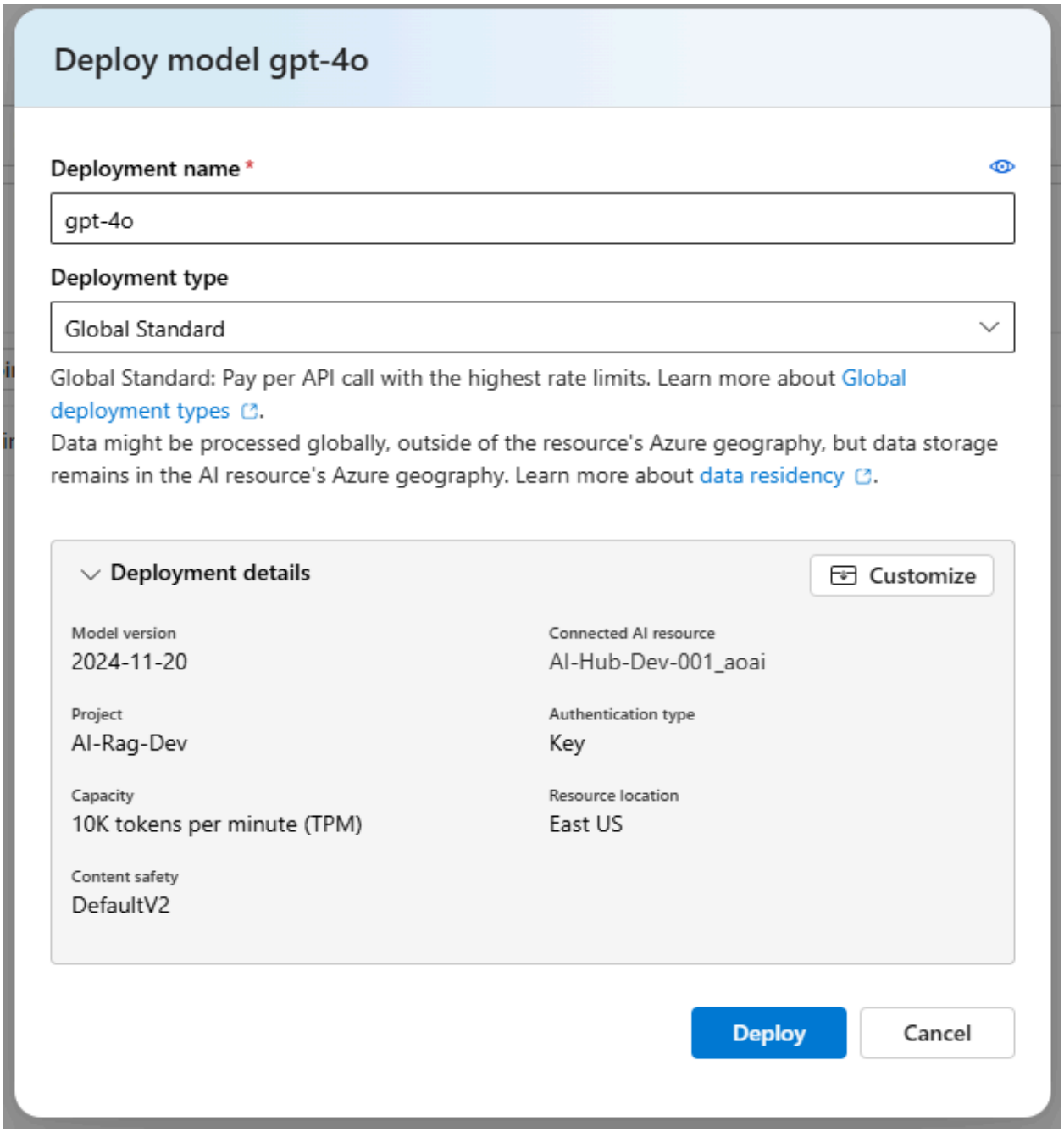
Create a new project under the hub created. Note regardless a project needs to have a Hub as a parent if one is not specified that one by default is created. Once the project is created and you navigate to the dashboard for the project you will see something similar to this. Note the project name in the top left “AI-Rag-Dev”.



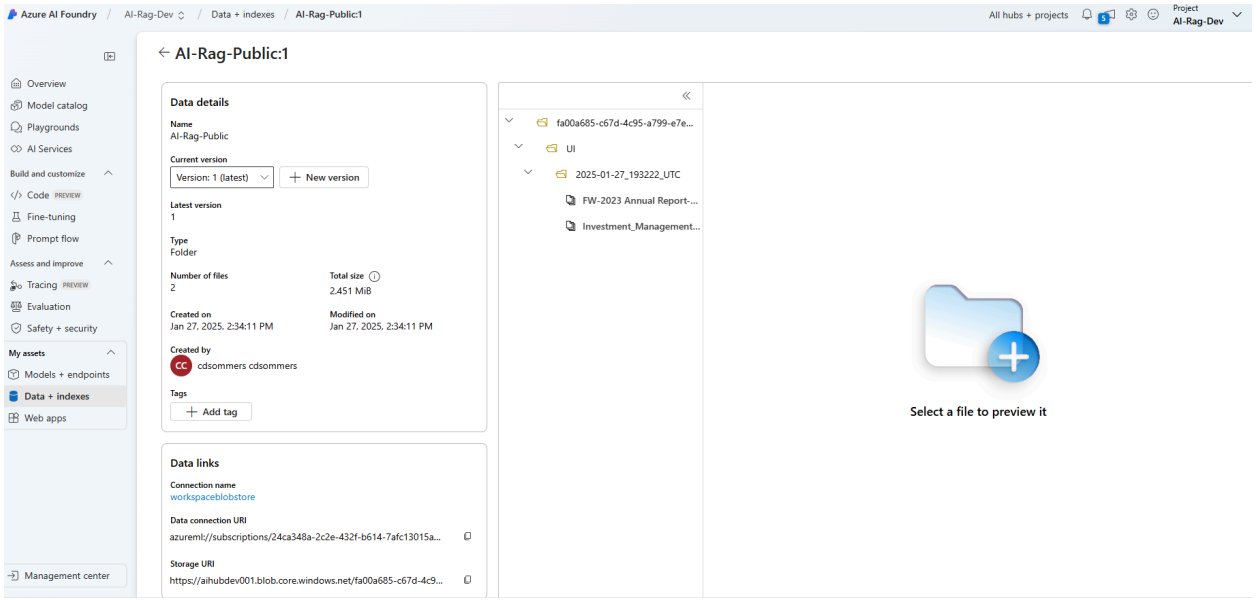
Deploy base model. Select Models + endpoints. Note you will need to deploy a model at allows function calling if you are using function calling in flow. I will deploy gpt-4o.



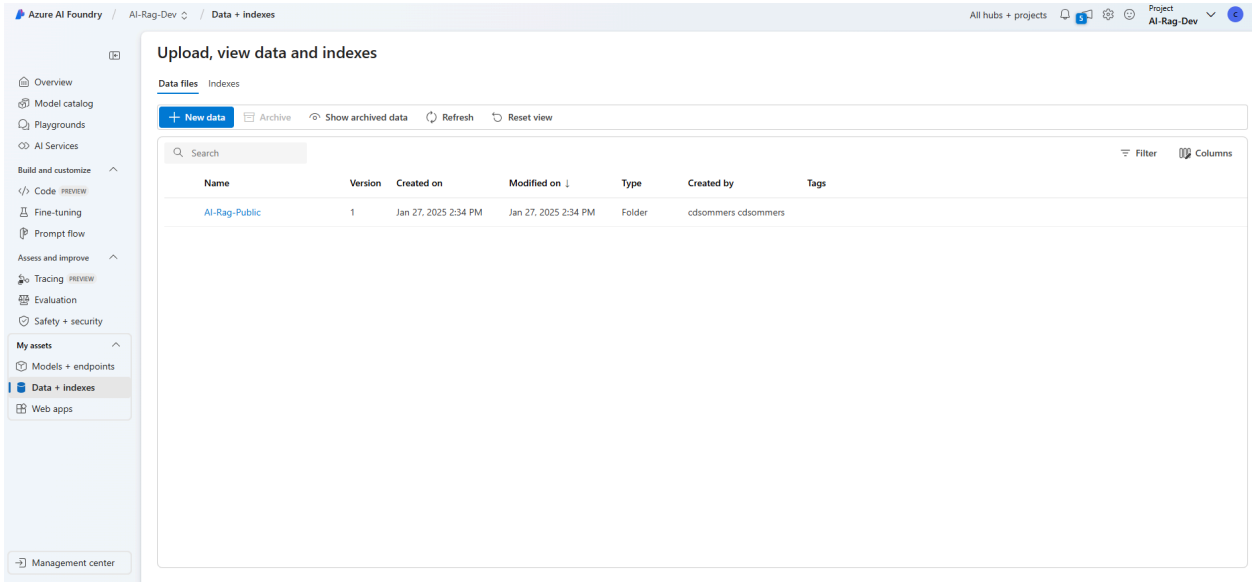
Select deployment type I am selecting Global Standard.



Create a data store. Click Date + indexes and walk through the wizard.

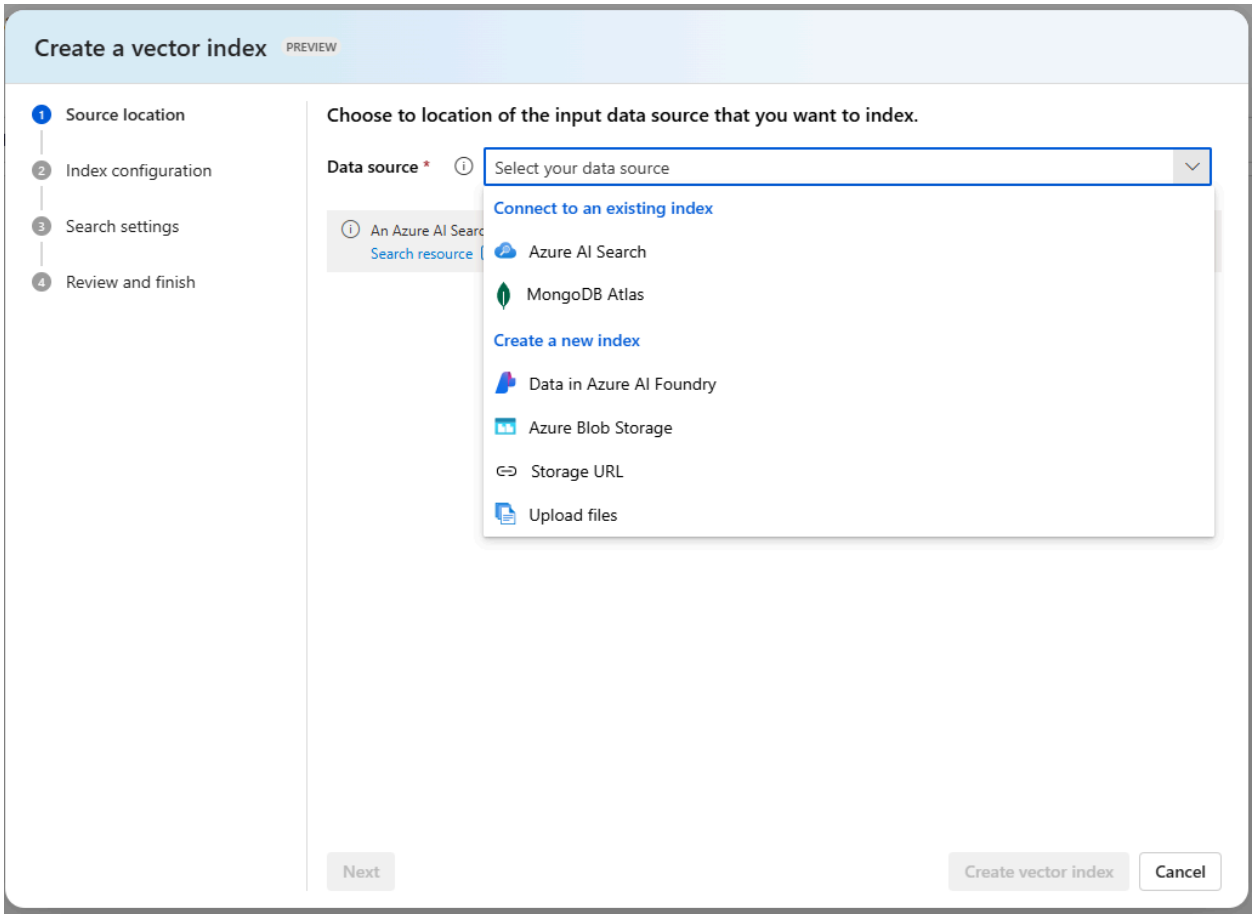


Note the connection name at the bottom left. The documents are stored within a blob storage.

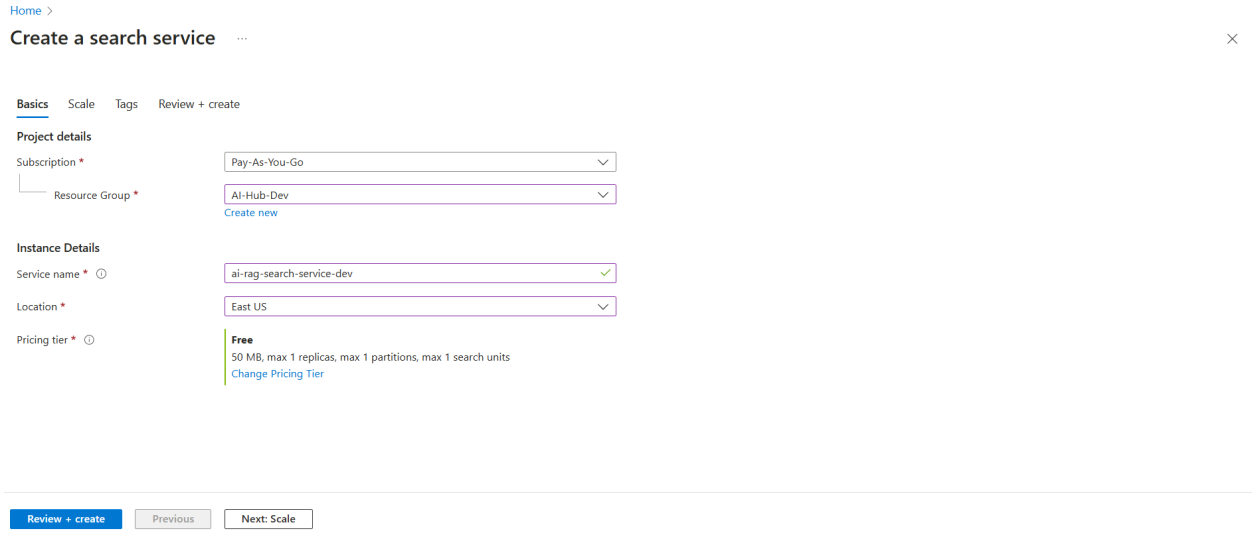


The Data store should be visible on the Data + indexes pane.

Create an index on the data. Click the “Indexes” tab on the Data + indexes pane. When creating a vector index click “Data in Azure AI foundry” to access the data stored in the data store in the previous step. (see below)

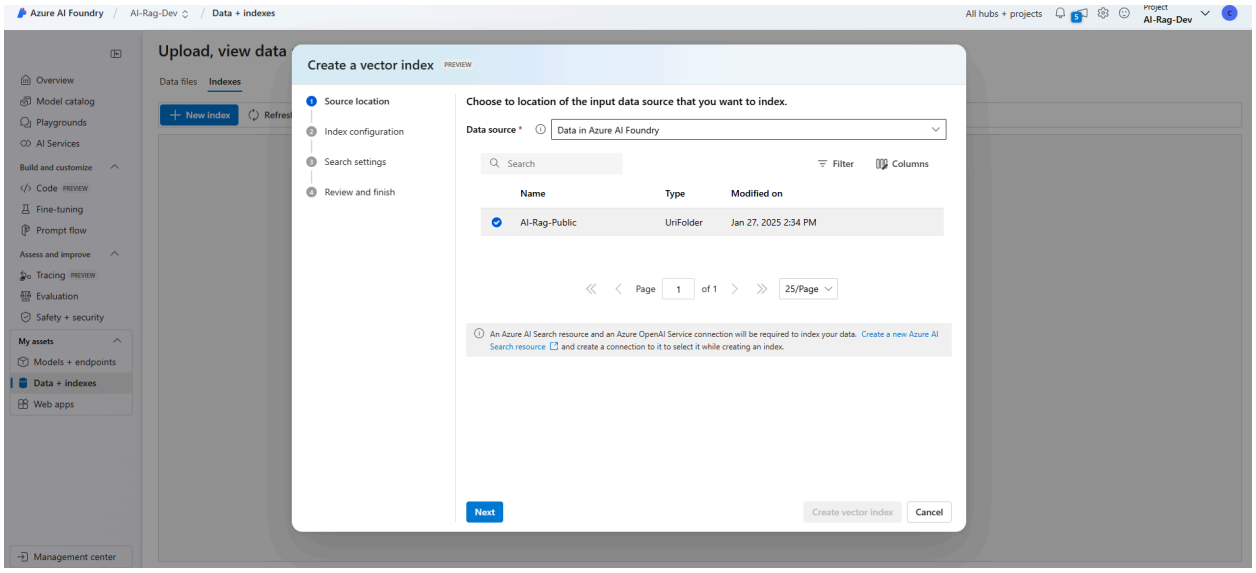


For index configuration you will need to create a new AI search service. This should open a new tab to create a new vector search service.

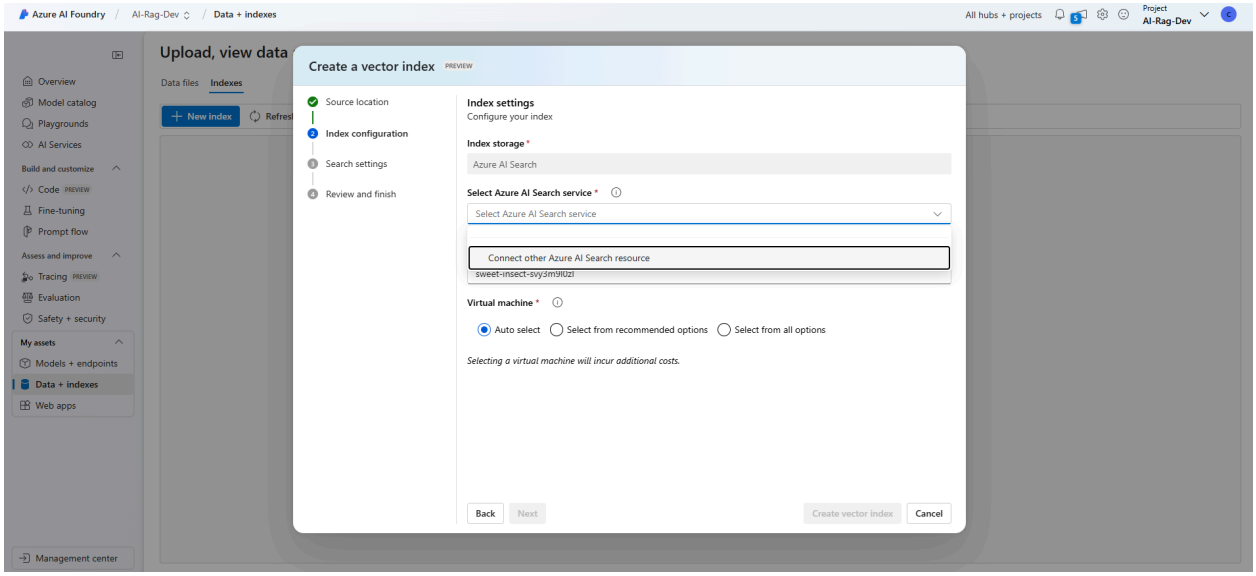


Note the pricing tier in this case I leverage a free tier for dev. Please take note of the pricing for each tier.

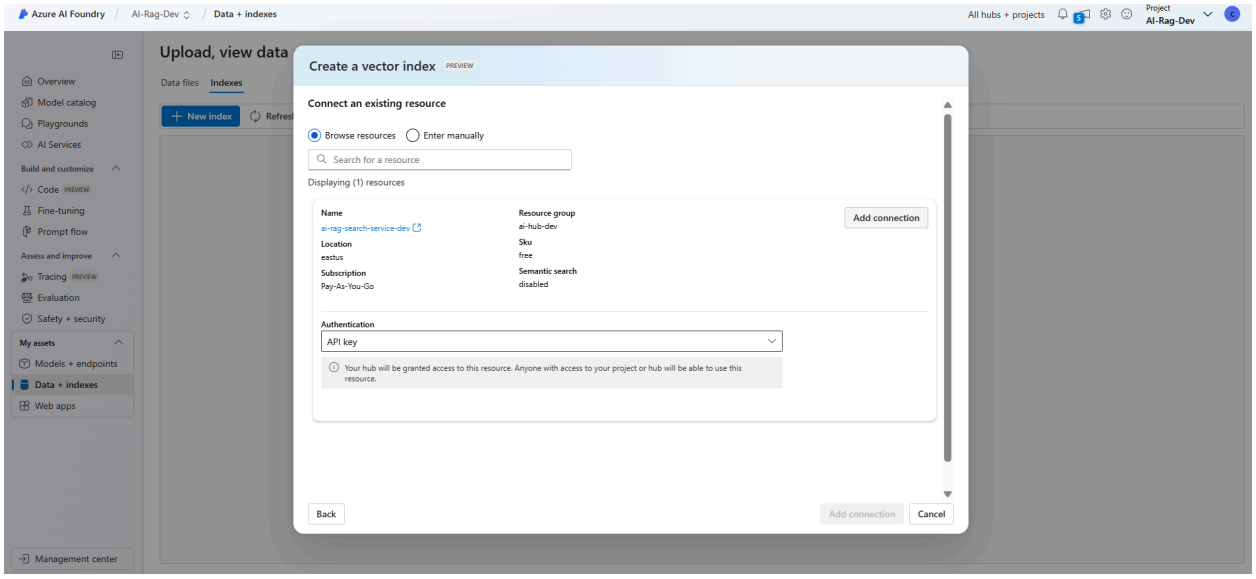
Once the search service is created create go back to the indexes tab and create index.
Select the data store in the previous step



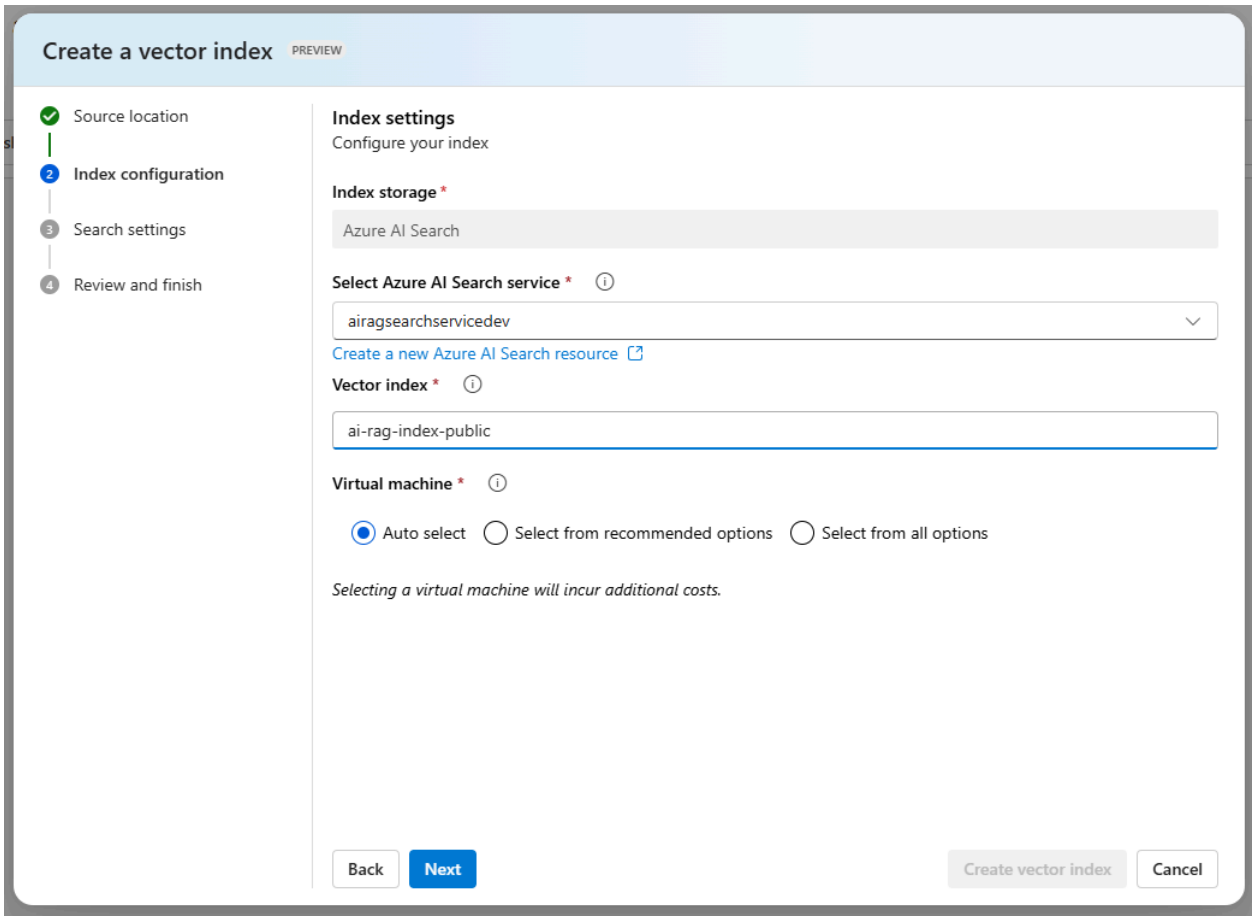
Select connect to azure search resource

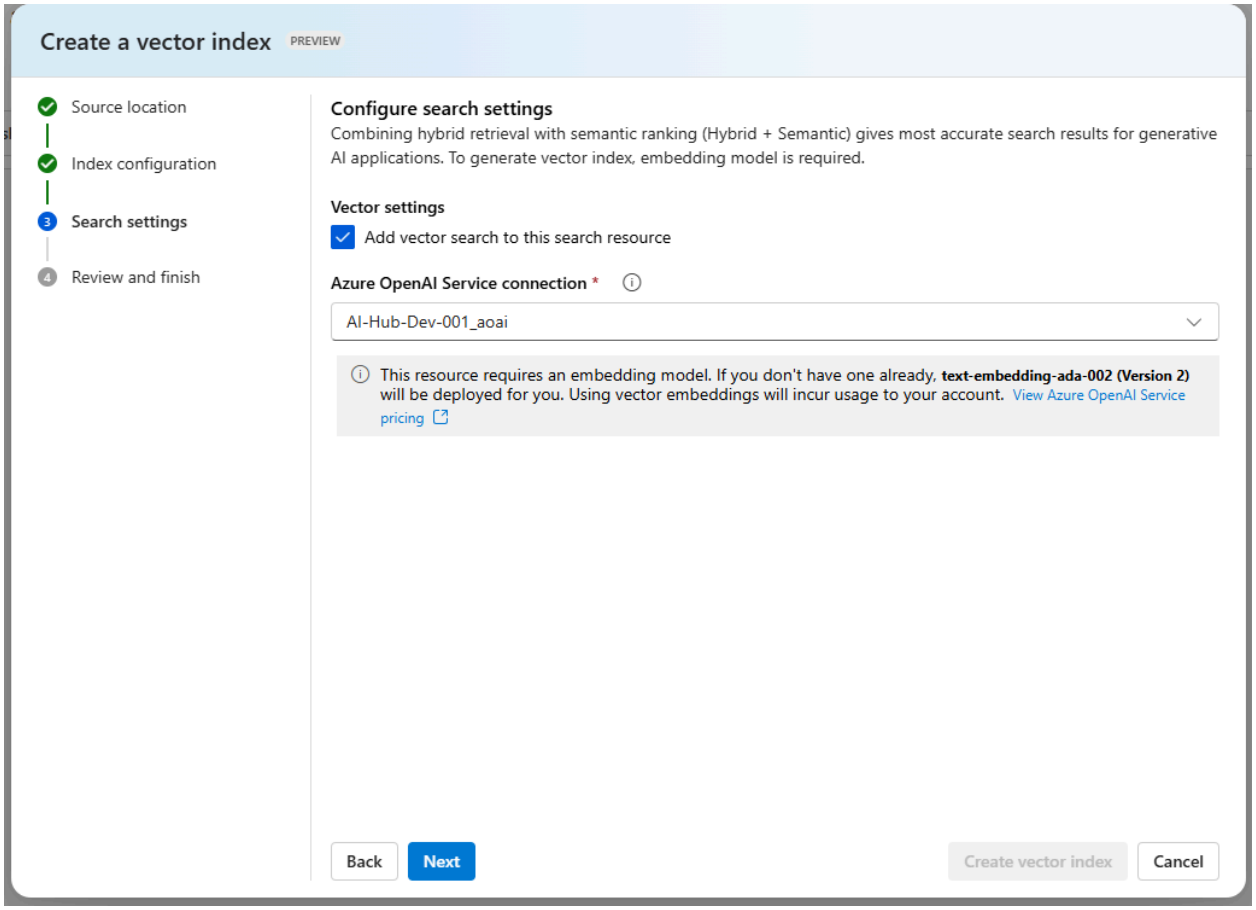


Select add connection to search service.

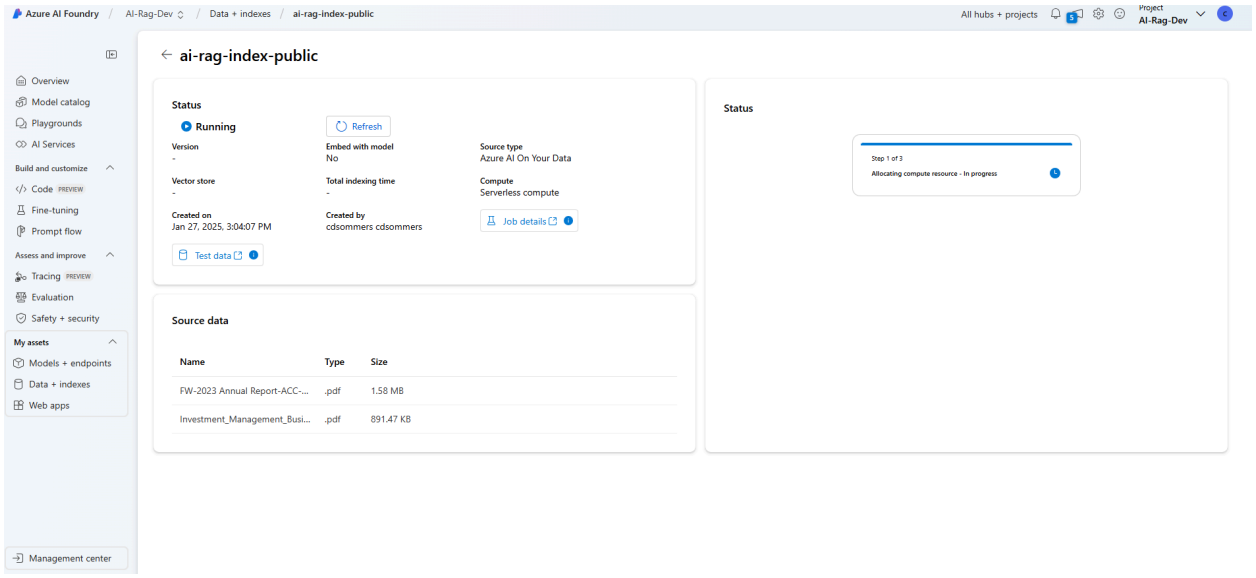


Give the vector index a logical name

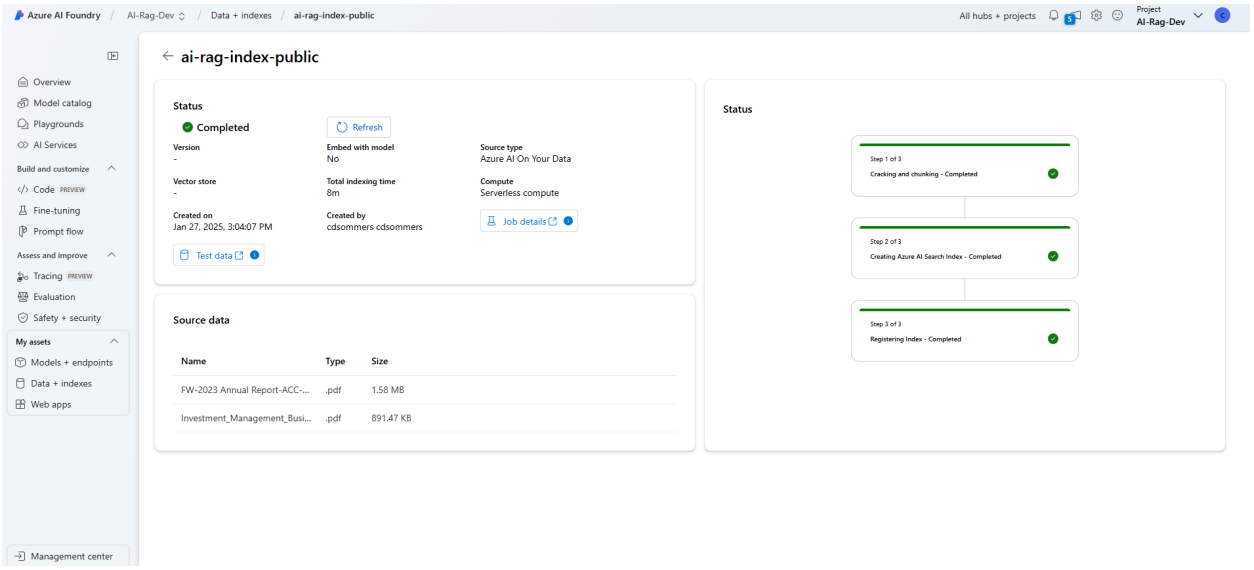


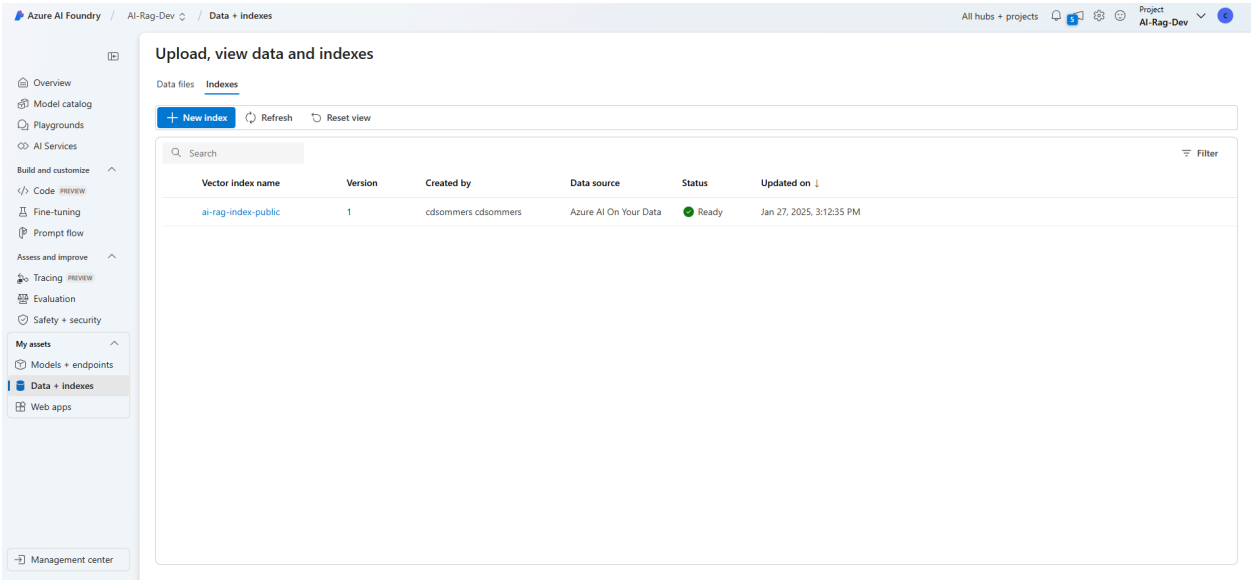


Click create vector index.



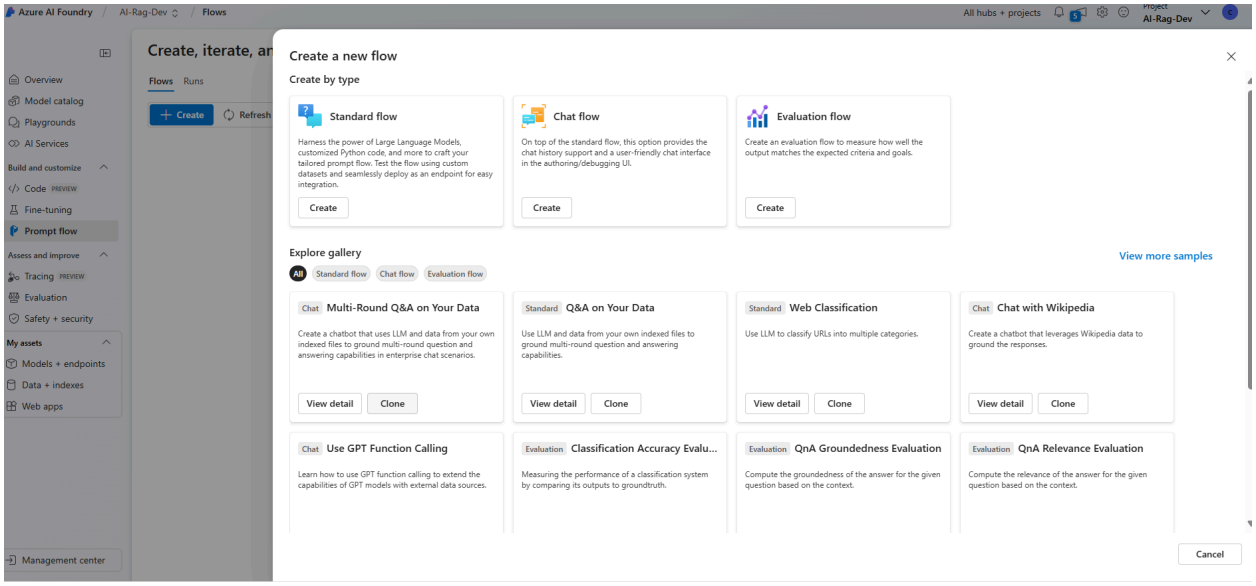
Vector index job should start. This should take some time.



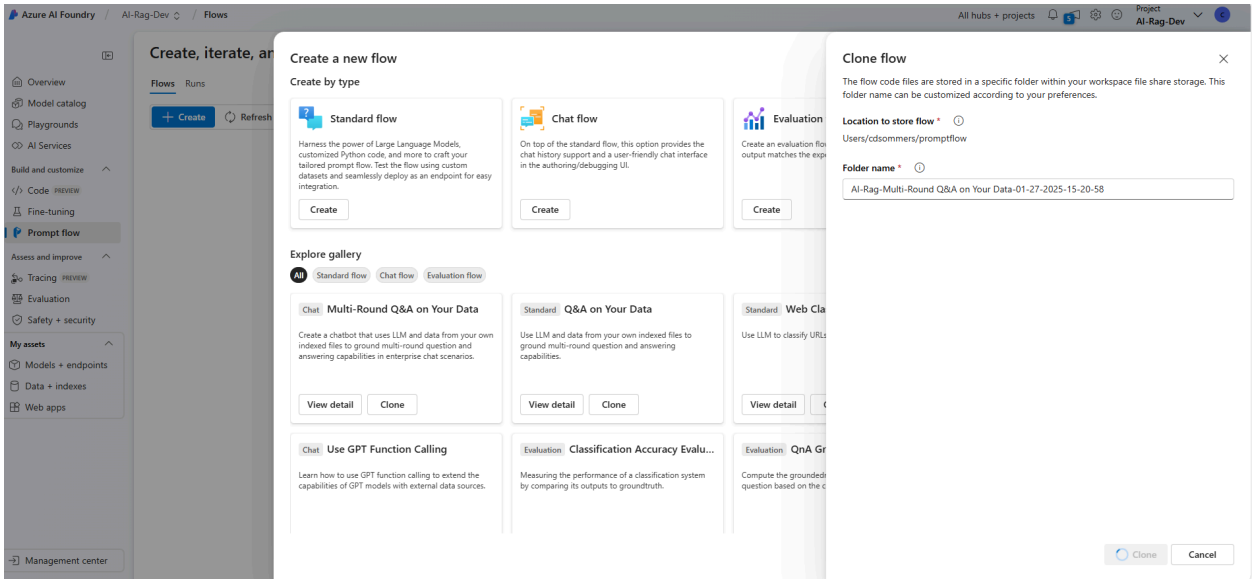


Once completed the index should have a status of ready.

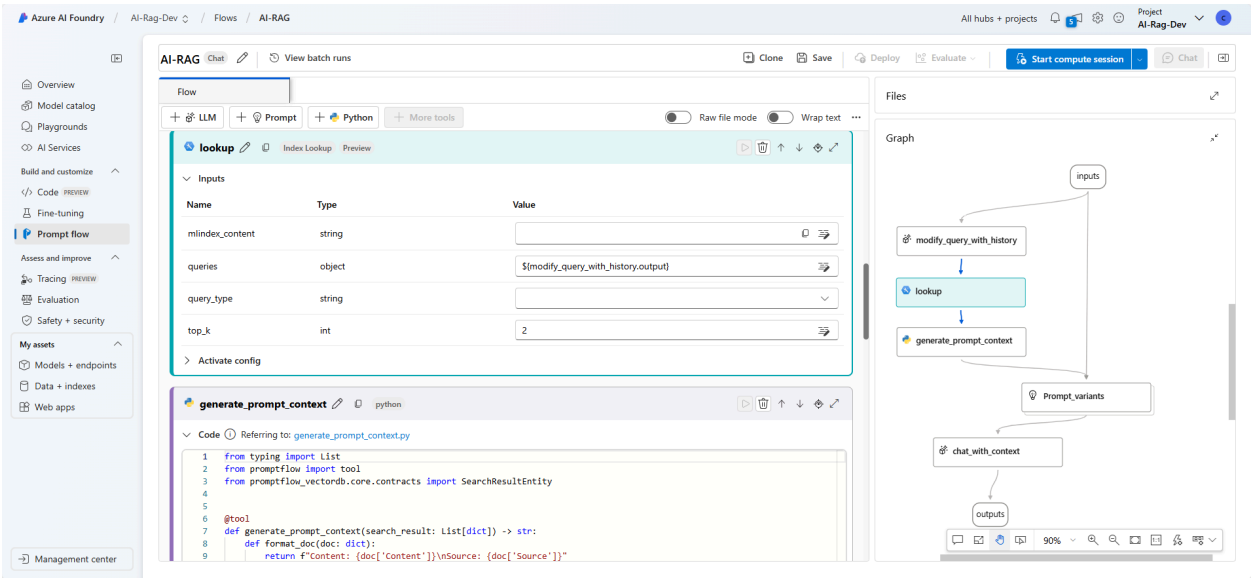
Create prompt flow. Create Prompt flow on left pane.
Clone “multi round Q&A on your data”



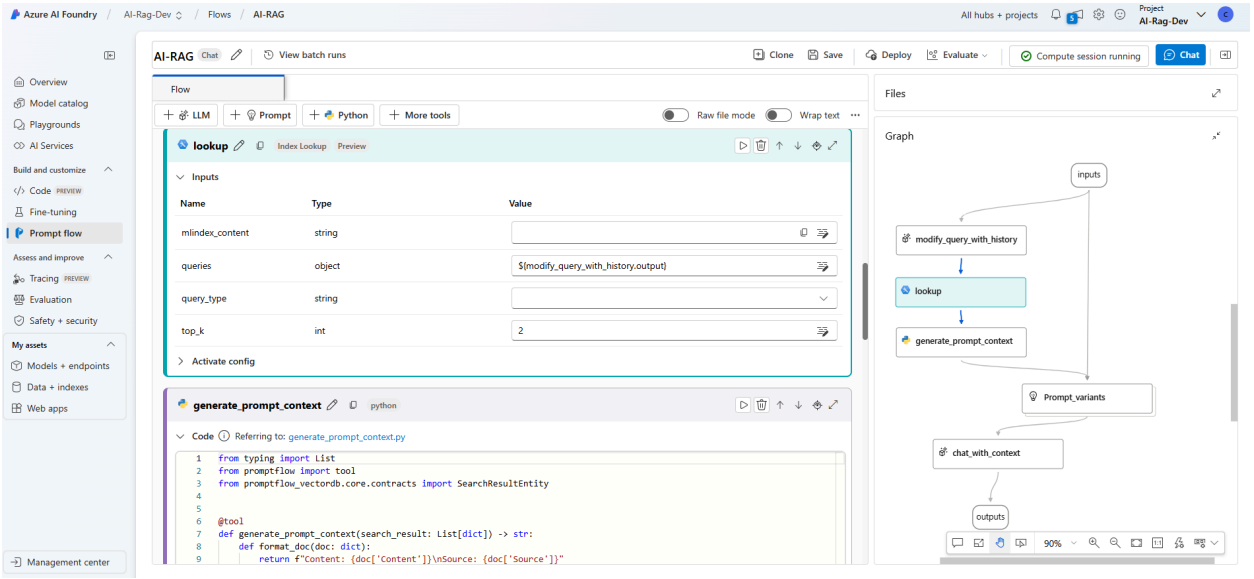
Create folder workspace



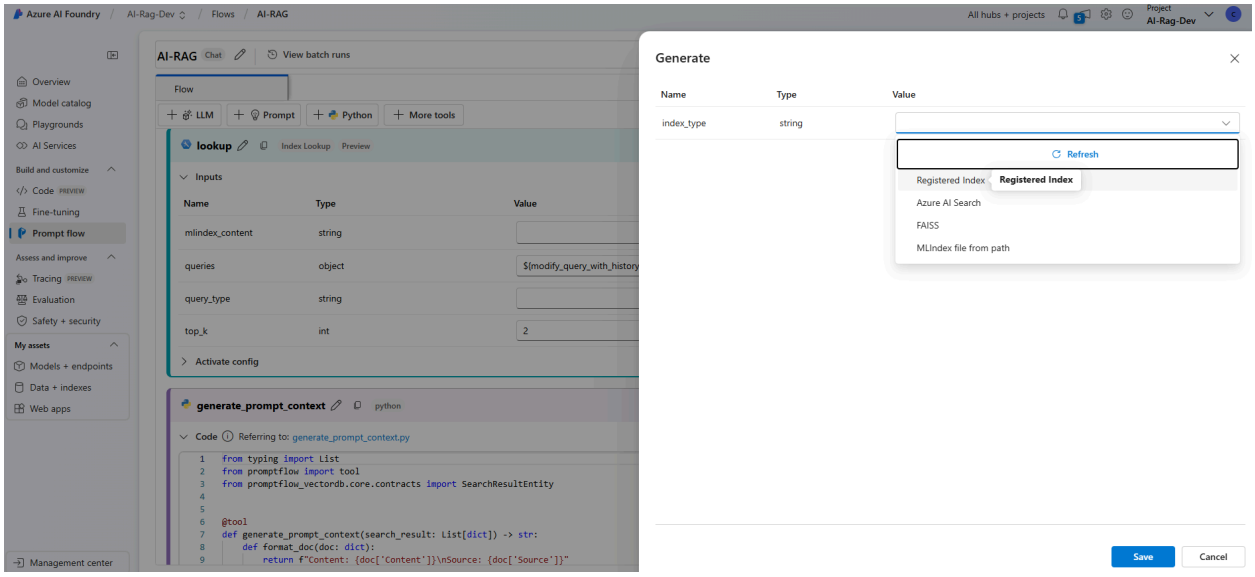
A new prompt flow should be generated



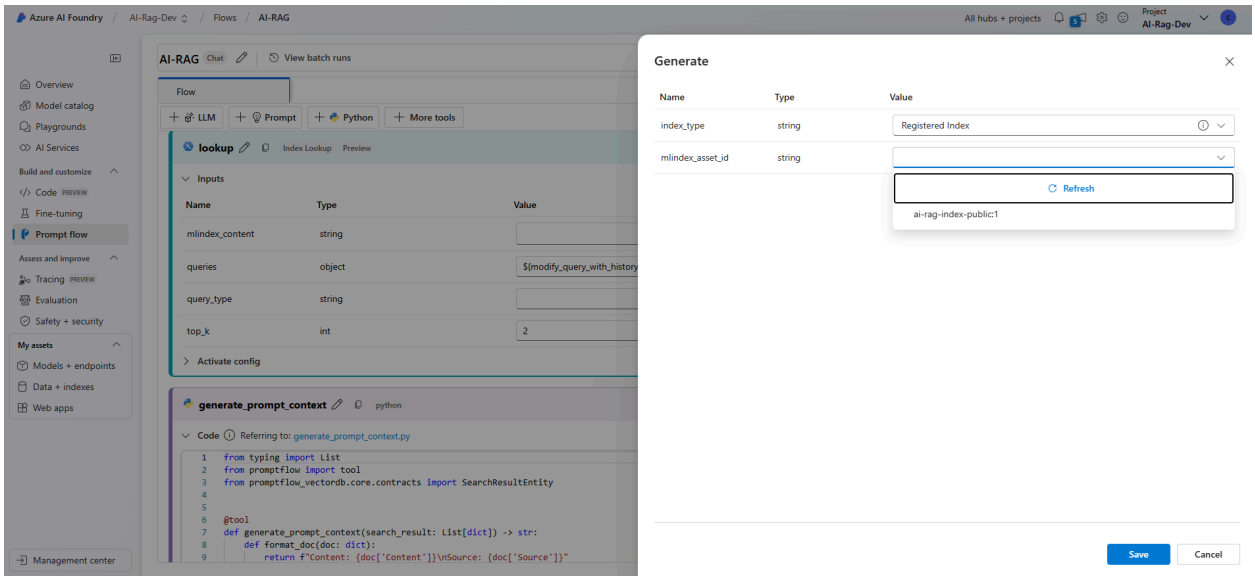
Start compute session



Select lookup node once session is started. Select mlindex_content and select registered index

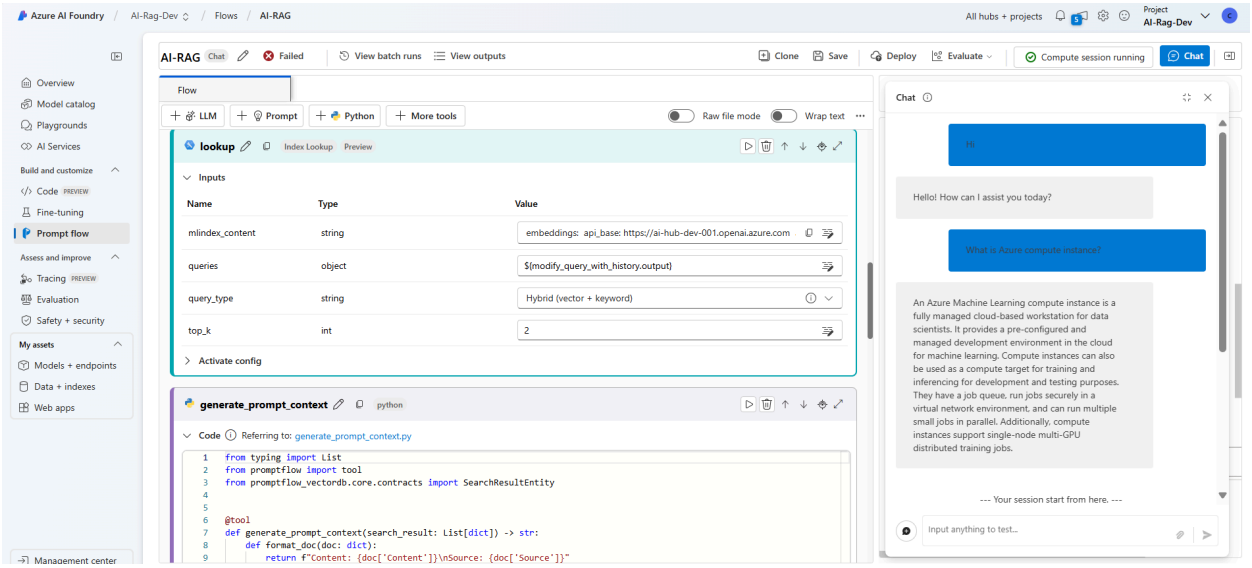


Select the index in previous steps.

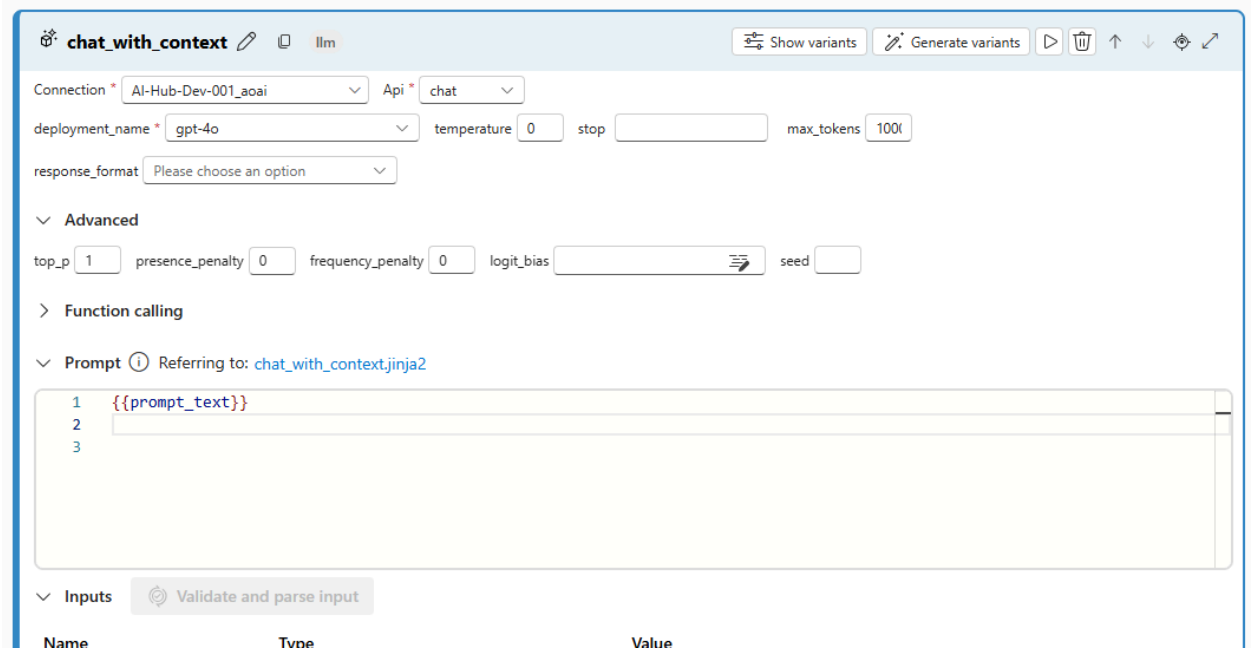


Select query type and select hybrid (vector + keyword)

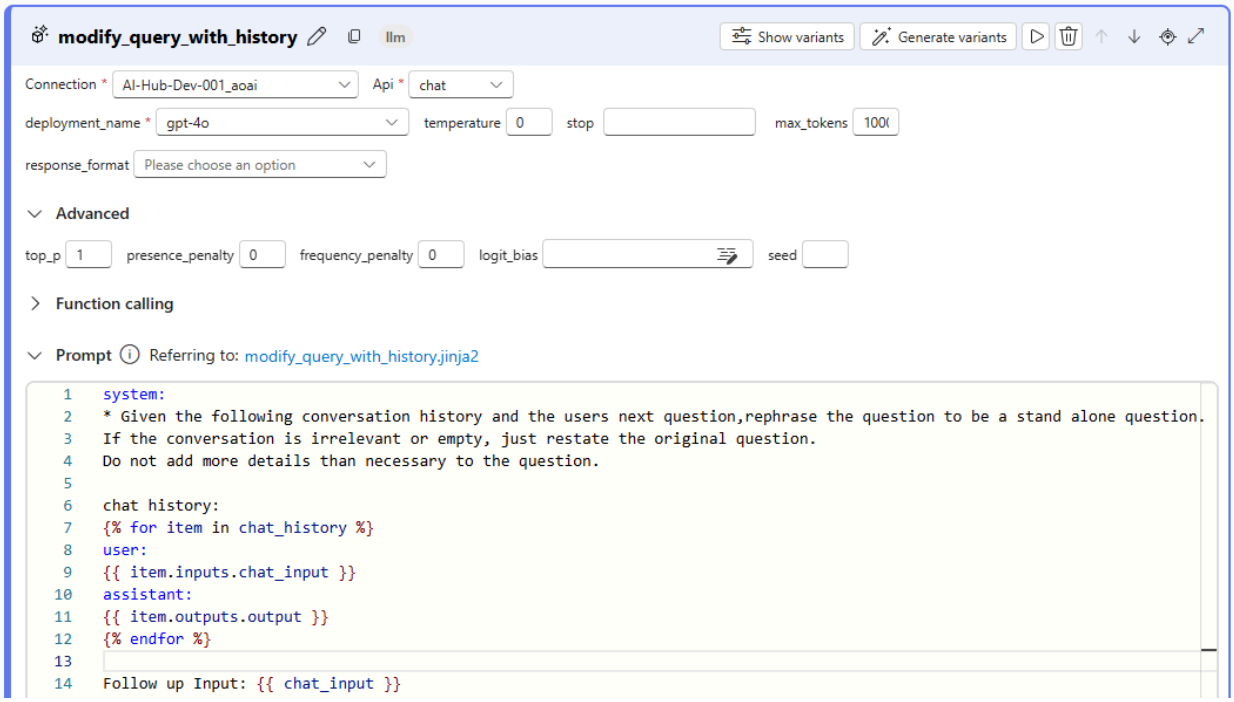
Click save



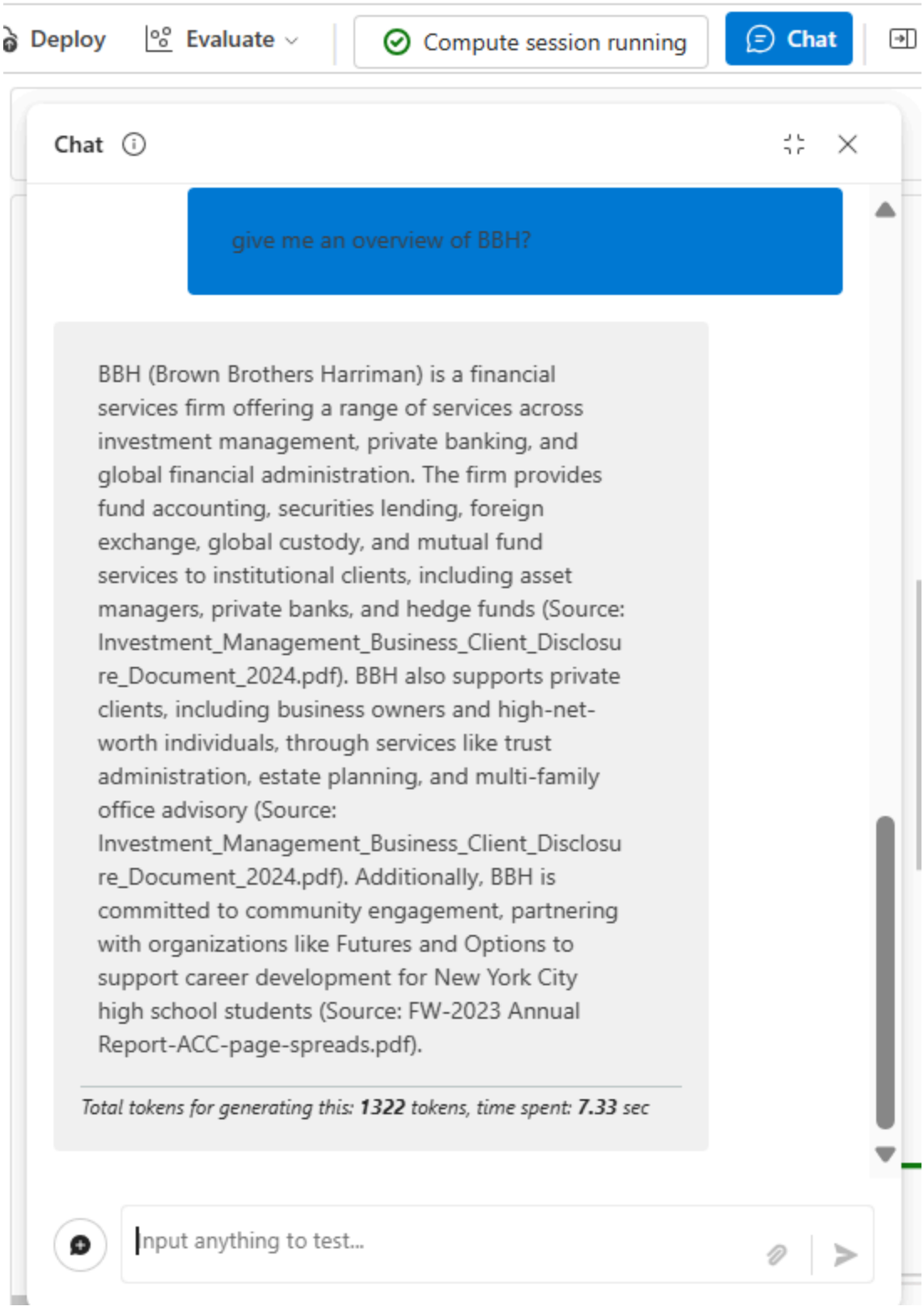
In chat with context step set the connection and deployment name (model)



In the modify_query_with_history step set the connection and deployment name (model)



Create question and click play to test


Deploy
Evaluate
Compute session running
Chat
Chat
give me an overview of BBH?
BBH (Brown Brothers Harriman) is a financial services firm offering a range of services across investment management, private banking, and global financial administration. The firm provides fund accounting, securities lending, foreign exchange, global custody, and mutual fund services to institutional clients, including asset managers, private banks, and hedge funds (Source: Investment_Management_Business_Client_Disclosure_Document_2024.pdf). BBH also supports private clients, including business owners and high-net-worth individuals, through services like trust administration, estate planning, and multi-family office advisory (Source: Investment_Management_Business_Client_Disclosure_Document_2024.pdf). Additionally, BBH is committed to community engagement, partnering with organizations like Futures and Options to support career development for New York City high school students (Source: FW-2023 Annual Report-ACC-page-spreads.pdf).
Total tokens for generating this: 1322 tokens, time spent: 7.33 sec
Input anything to test...