Build your own virtual assistant using genarative AI

In this lab we will go over how to set up a virtual assistant (or chatbot) for customers using IBM Watson services: Watsonx.ai, Watsonx Assistant, and Watson Discovery.

We created a Retrieval Augmented Generation (RAG) pattern to answer user questions with Large Language Models (LLM) using the existing documentation which resides inside the knowledge base.

Functionality performed by the three Watson components:

- Watson Discovery - The knowledge base that contains all the data (web crawled, PDFs, etc.) which the chatbot will base it's answers on.
- Watsonx Assistant - The conversation service through which the user interacts with the chatbot.
- Watsonx.ai - The service that will call the LLM, enabling virtual assistant to generate responses based on user questions.

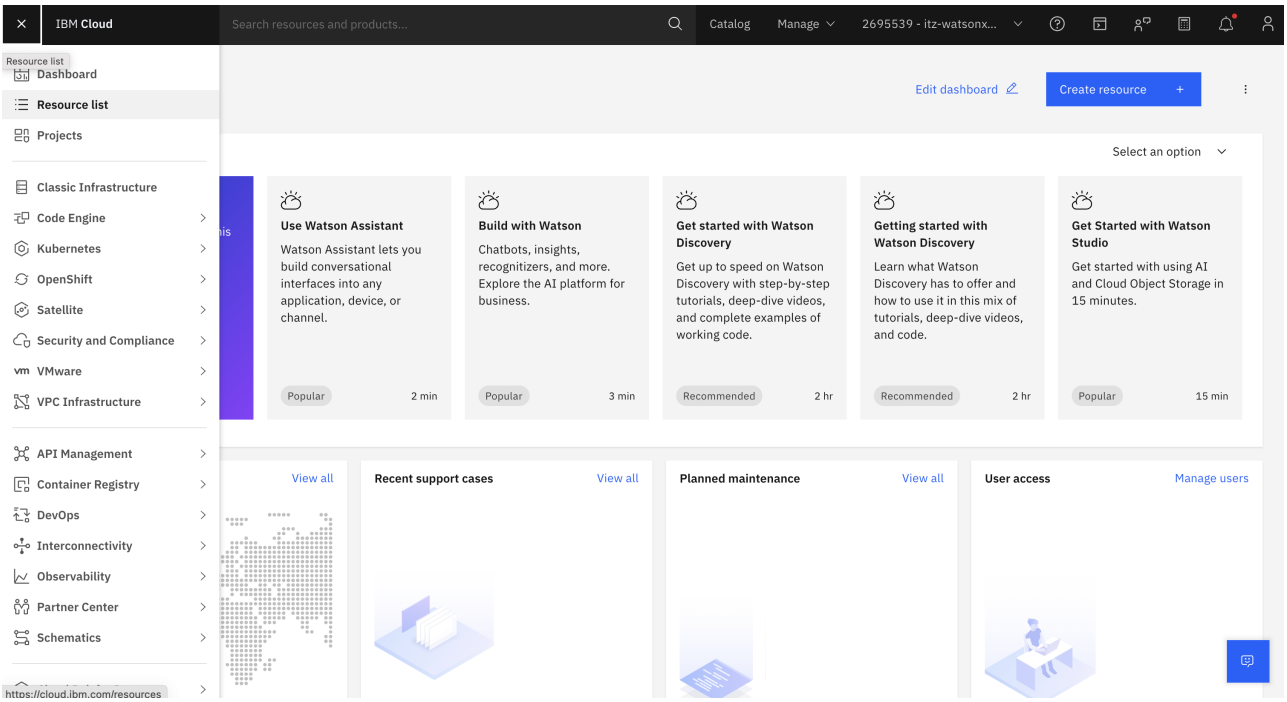
1. Set up Watson Discovery and database

In this chapter we will go through how to set up the Watson Discovery and create a database from different sources.

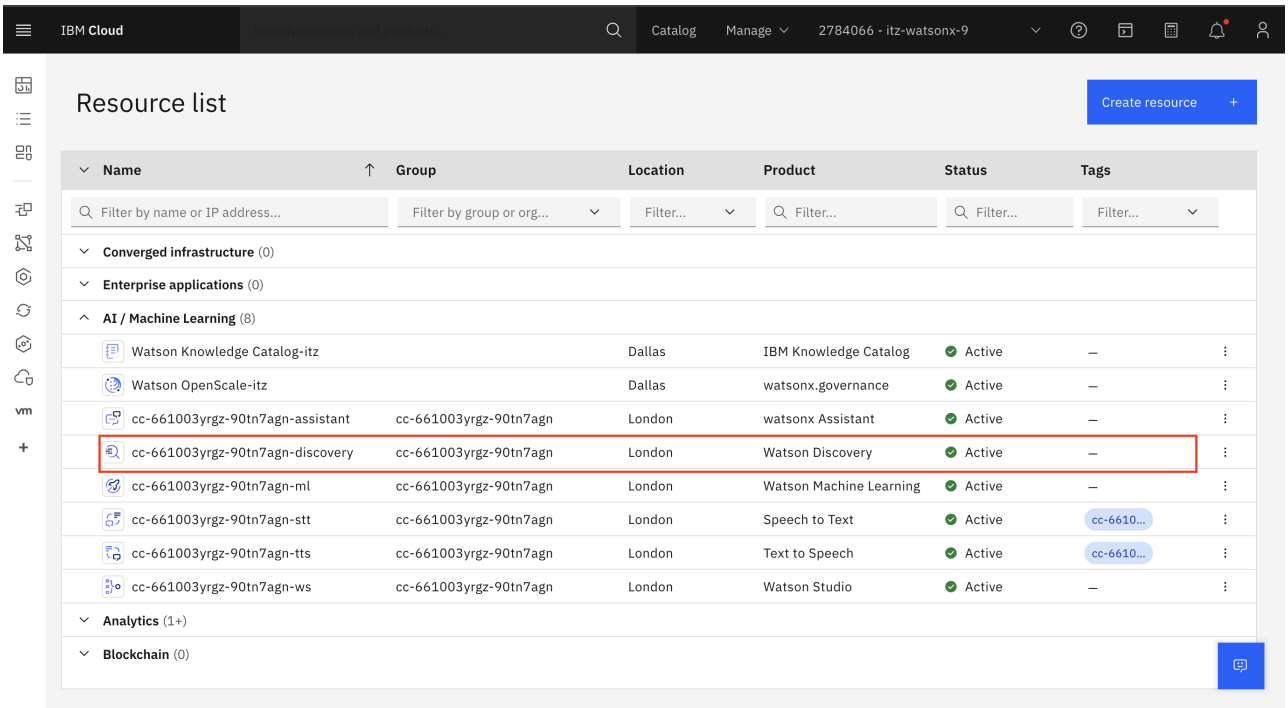
1.1. Creating a Watson Discovery project

Before we can config the database we need to create a project in Watson Discovery where we will store the data, to do so follow these steps:

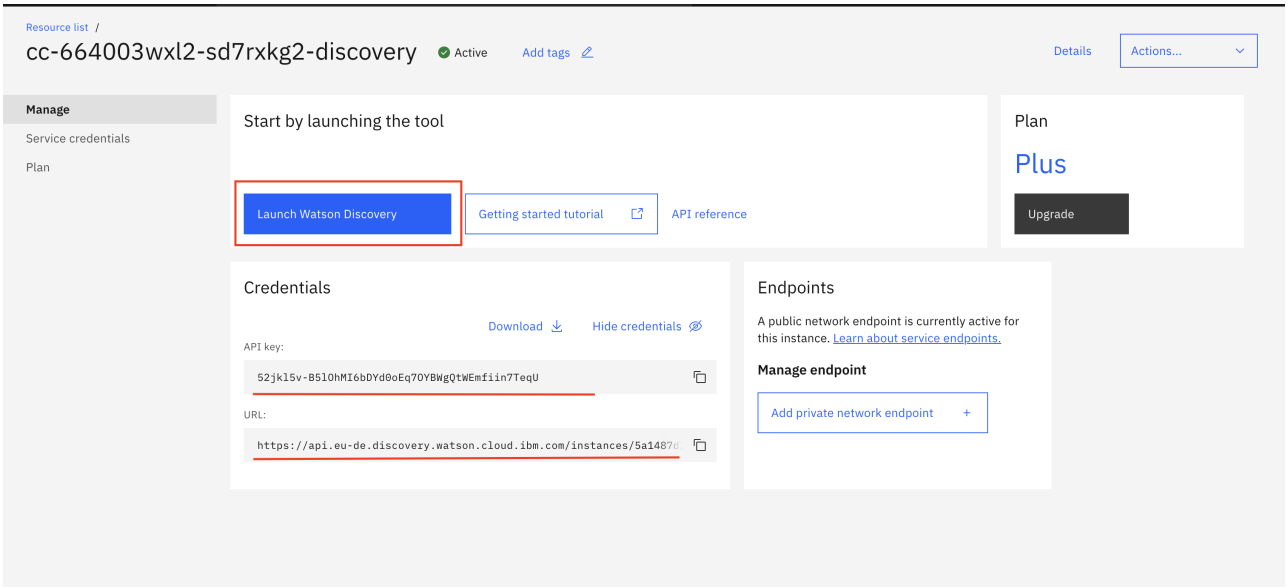
1. Open the resource list



2. Click on Watson Discovery in the resource list



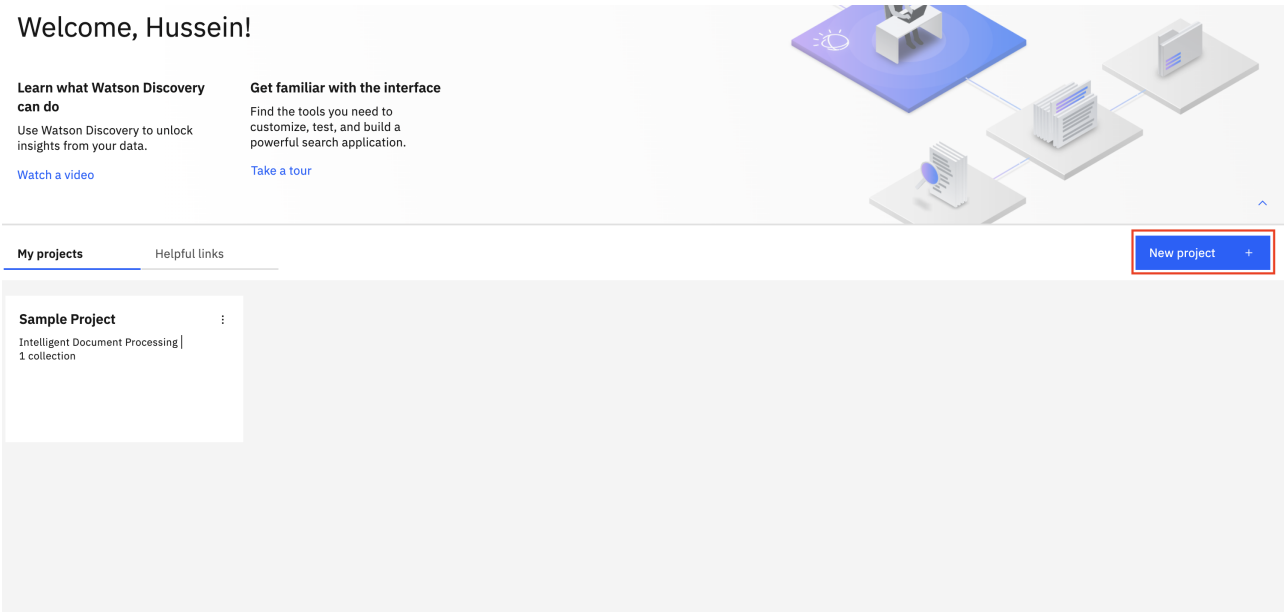
3. Click "Launch Watson Discovery".



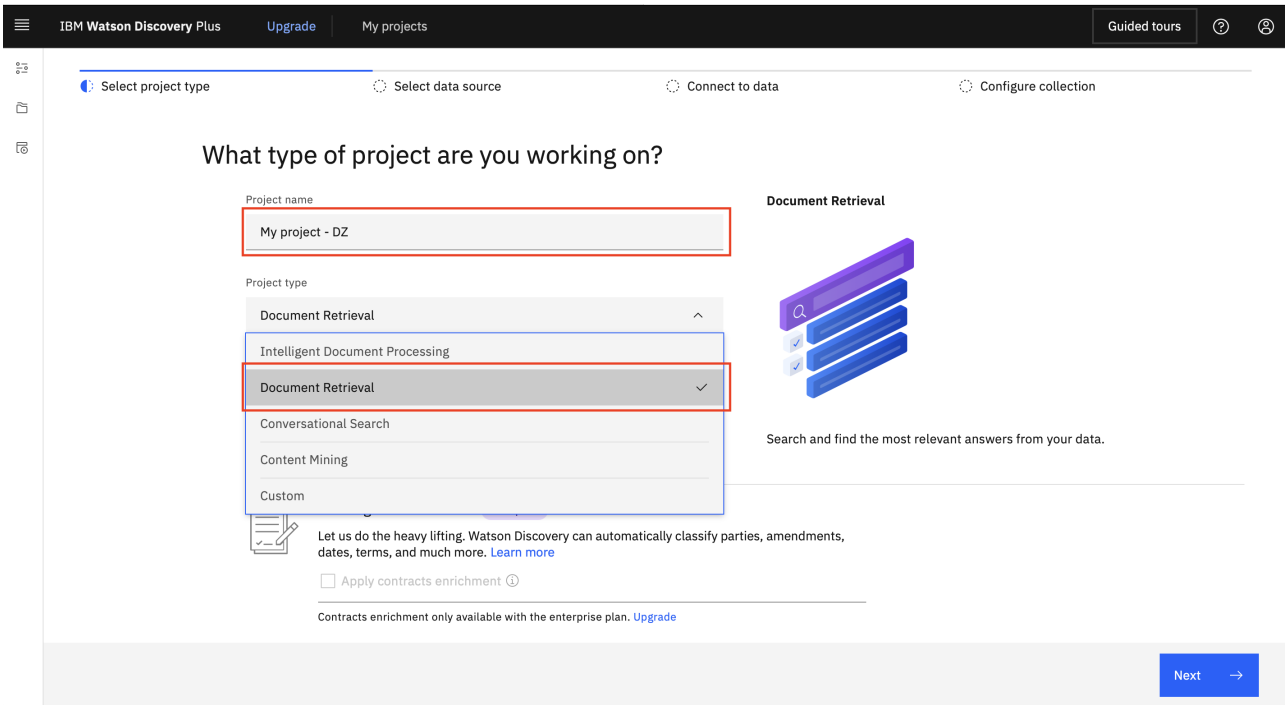
Note: The prefix of the resource name might be different than on the image.

Note: We will use the credentials from Watson Discovery later so save the **API key** and **URL** in a document.

4. Click on the **New Project** button



5. Give it a **name with your initial** so you can always find your project. We are conducting the lab in a shared Watson Discovery instance and you may see others project. Make sure you always do changes in your project.



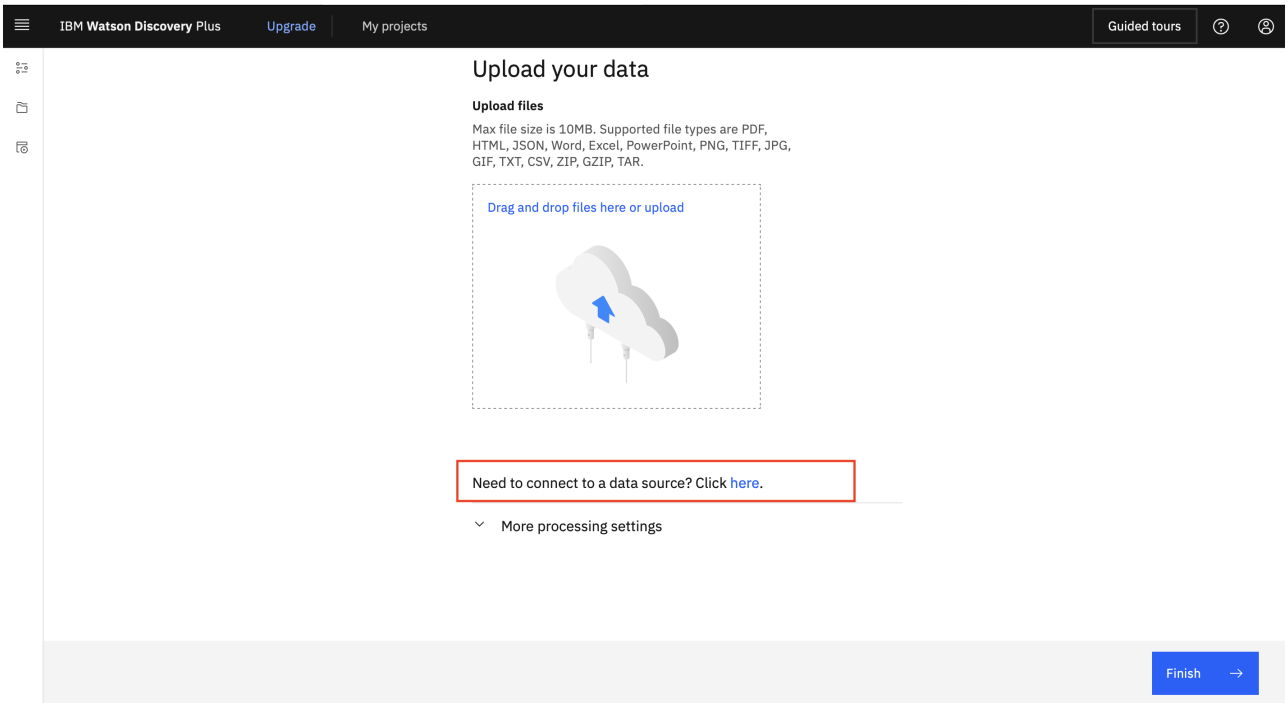
6. Select **document retrieval** and click next

1.2. Create Watson Discovery Collection

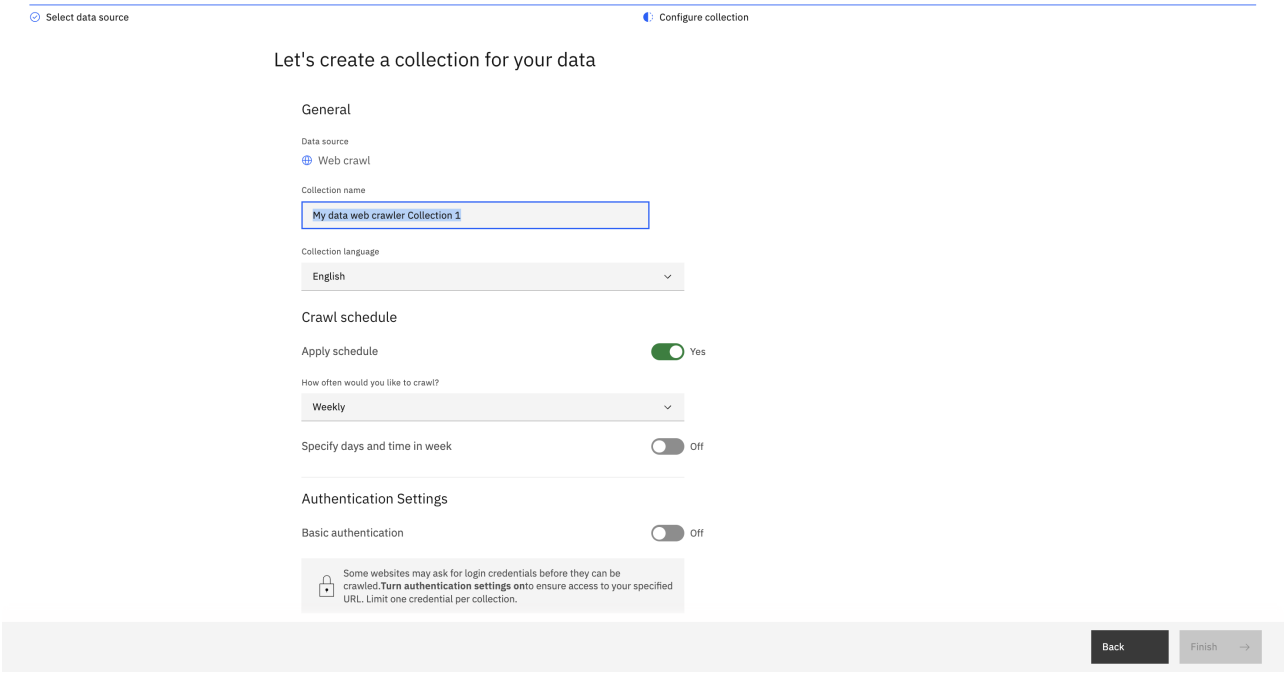
Now that you have created a project we need to create a collection where you can store your data. A collection is like a folder where you can store different data from different data sources in multiple collections inside one project.

1.2.1. Create a web crawling collection

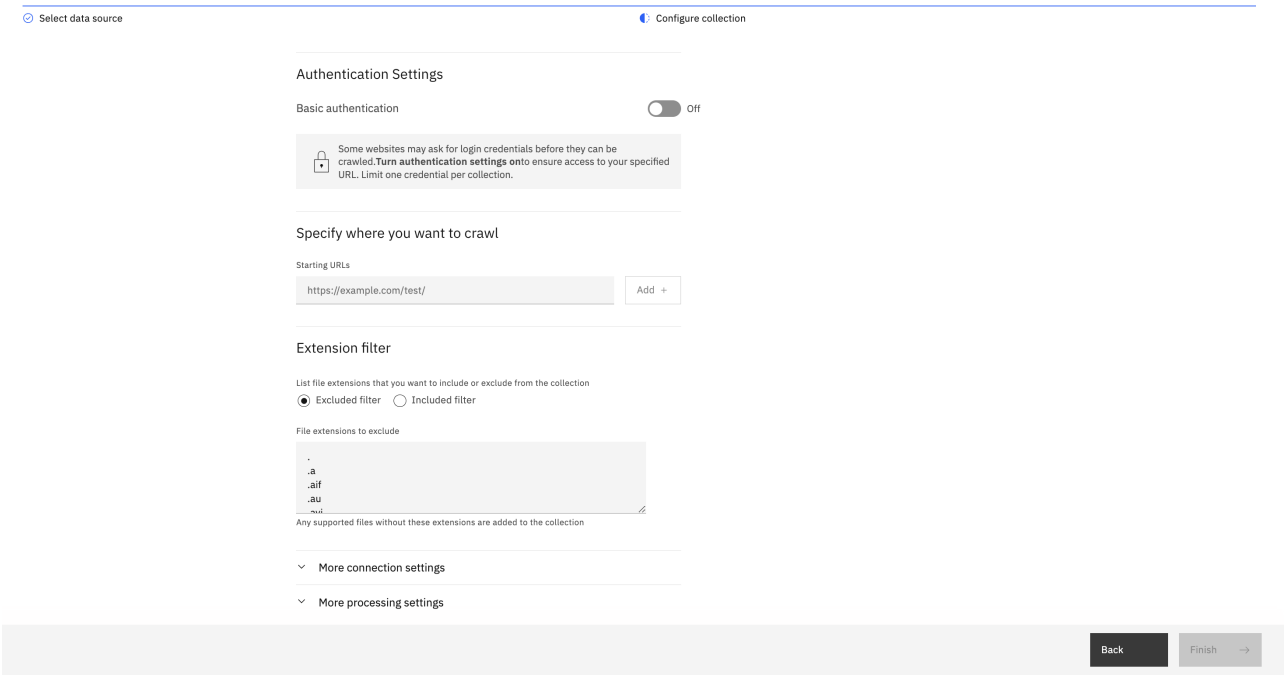
1. Scroll down the page and click the **here** text to connect to a data source



2. Choose the language of the data you want to crawl



3. Scroll down in the page and add the URL of the website that you would like to crawl, you can add multiple sections of a website. If you're unsure, call on the lab assistant. Press add when you're done.

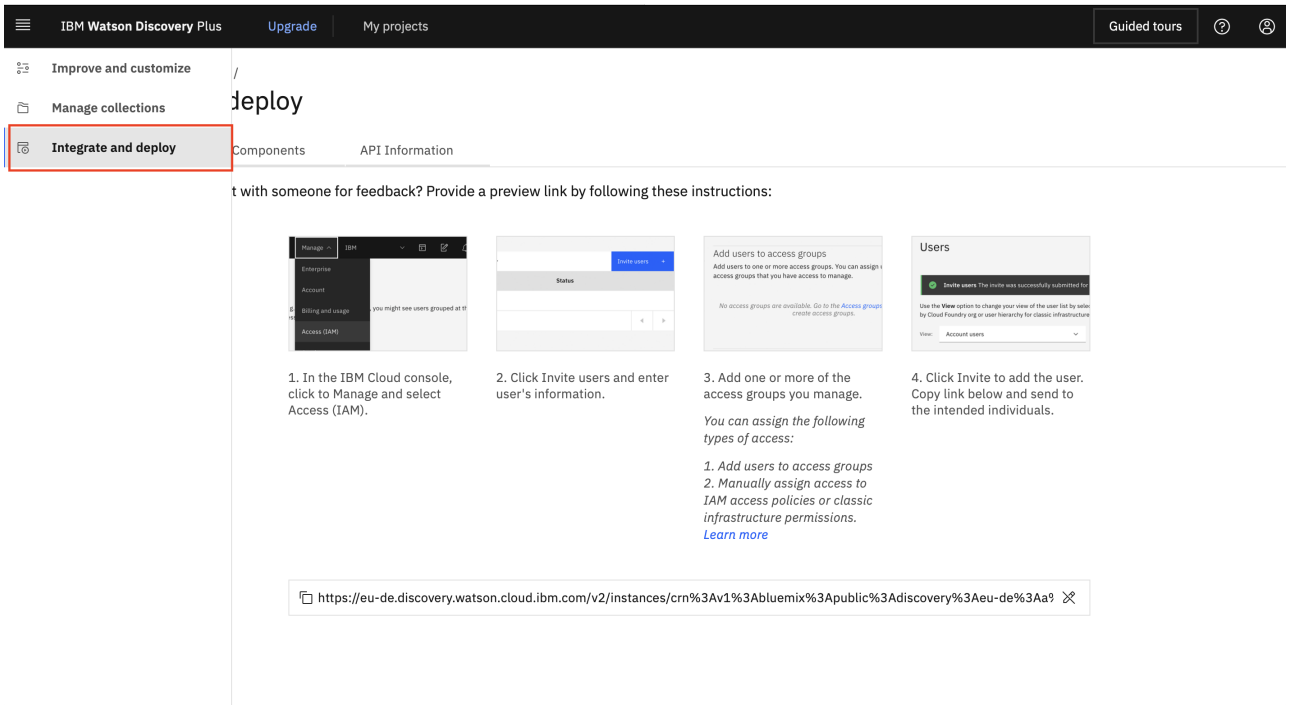


4. Click finish

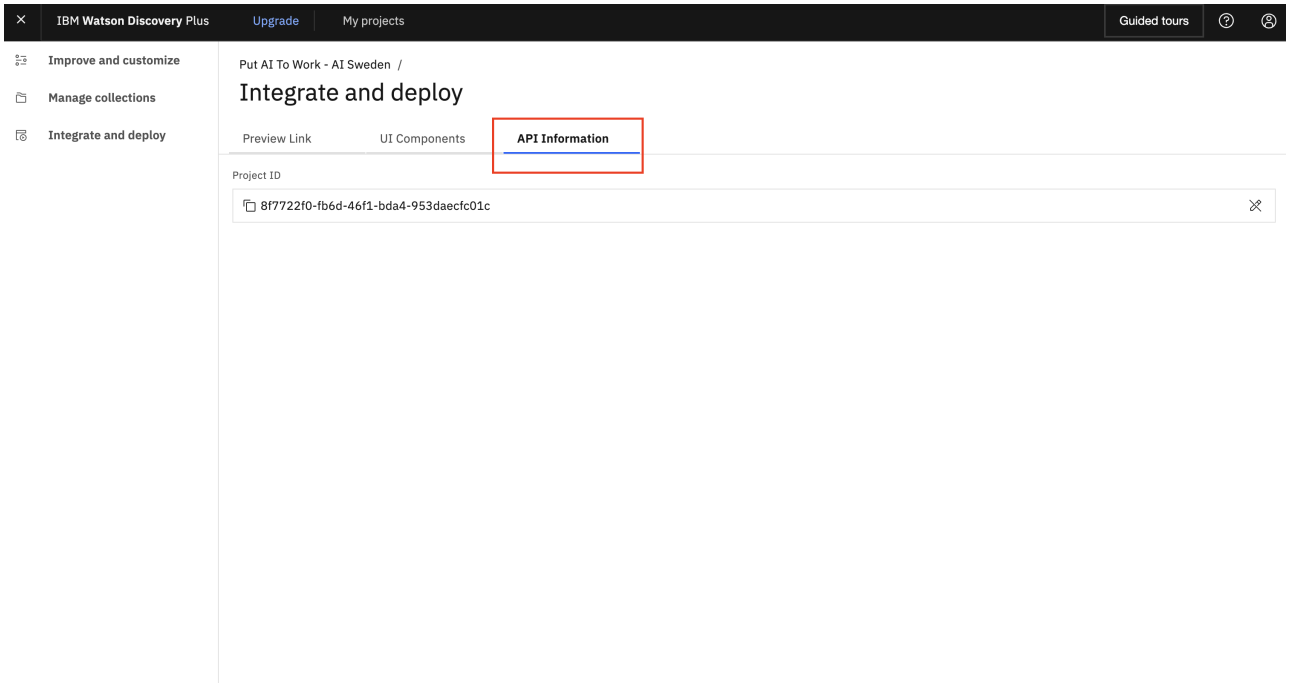
Watson Discovery will now start to prepare your data and it can takes some time. Meanwhile we can continue to set up other services.

1.3. Getting the project ID of Watson Discovery

1. On the left menu bar click on **Integrate and deploy**



2. Click on **API Information**



3. Save the **Project ID** in a document and mark with **WD-Project-ID**, we will use it later to integrate with Watsonx Assistant.

2. Watsonx.ai configuration

2.1. Credentials to Watsonx.ai

Watsonx.ai is the AI platform facillitate access to LLM. We have created a watsonx.ai project for you as well as the associated API key to the service.

Save the following credentials into your notepad. They will be used in the next steps where we integrate watsonx.ai the the chatbot.

> WX-API-key: GpId1X3uWWyZJYBDhkoT5u83heCDKt23Kx8sFMf0evKj
> WX-project-ID: df9e9ba2-128f-4f5e-a34e-ff5a21894136

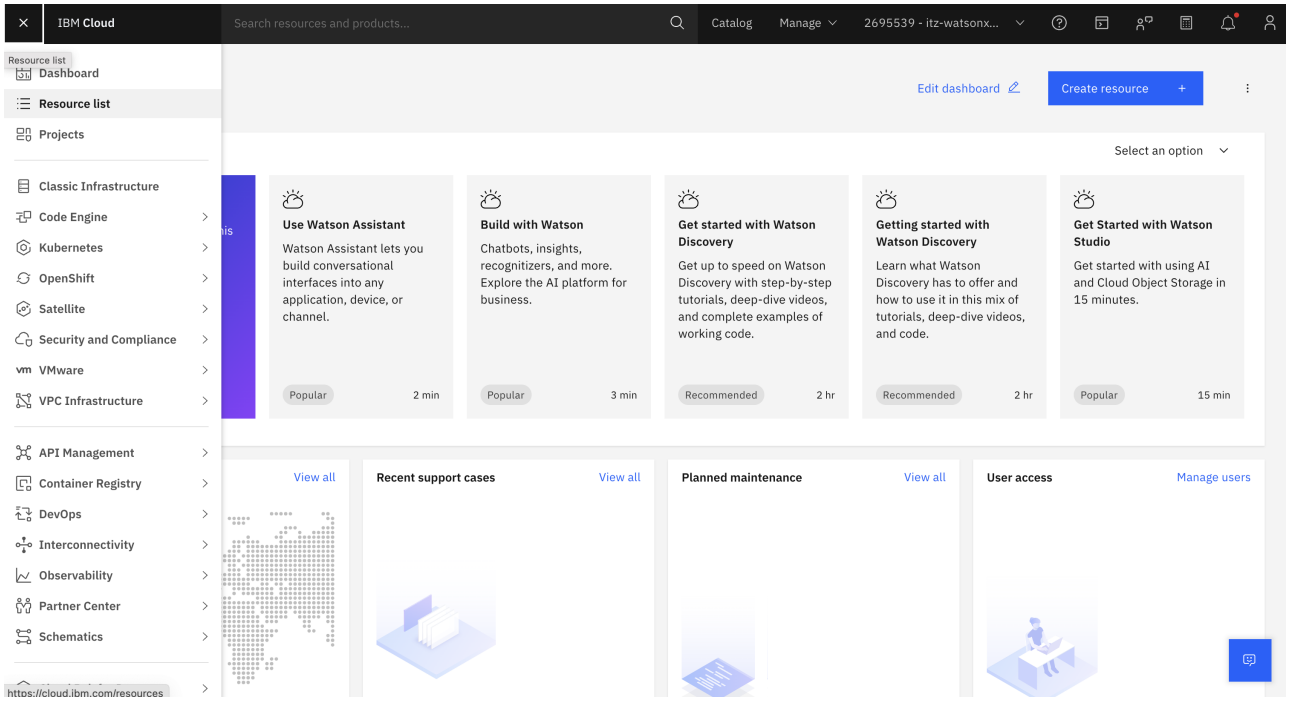
We have all the information we need. We can go to Watson Assistant to start creating the demo.

3. Setting up the Watsonx Assistant

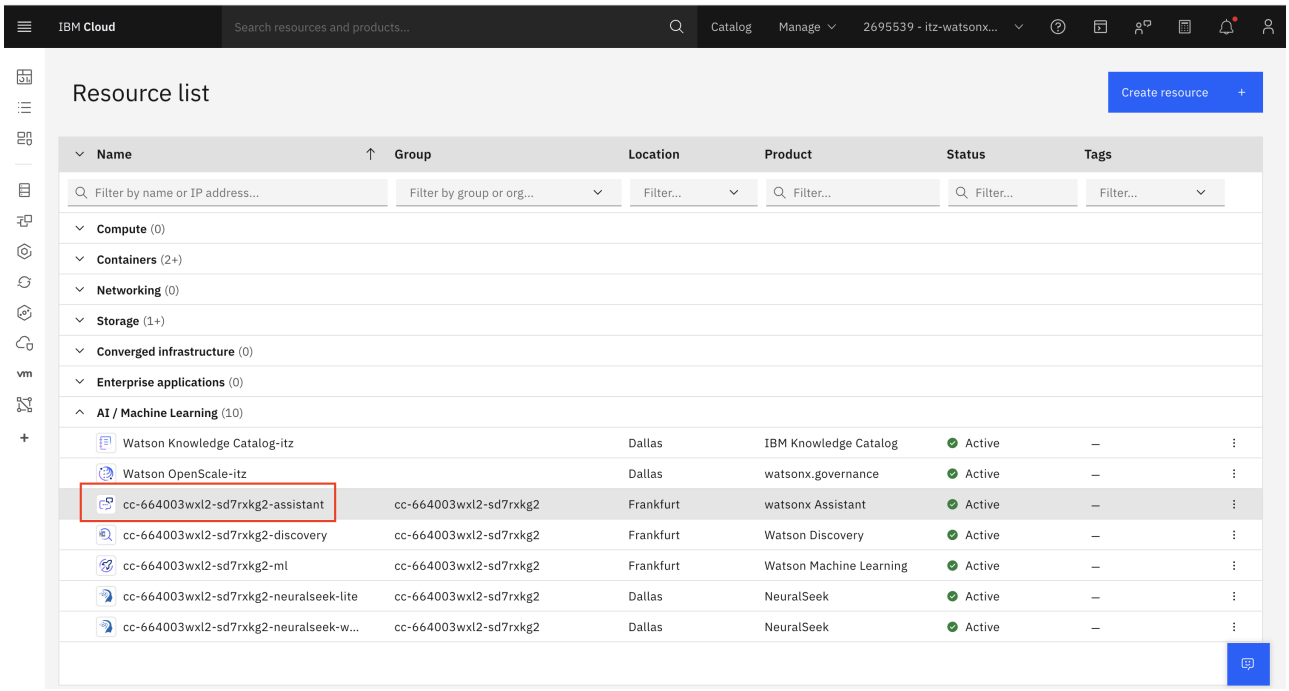
In this section we will integrate the other services so that we can ask the assistant questions.

3.1. Creating the a chatbot

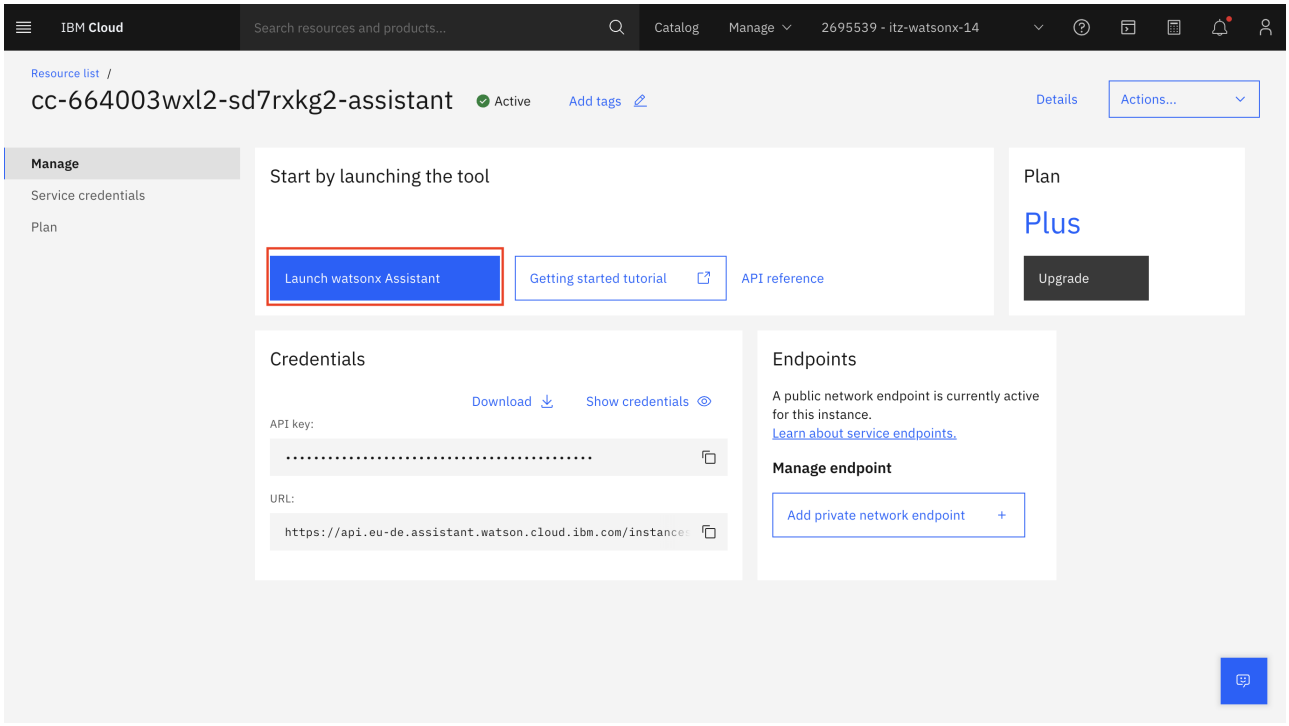
1. Go back to "IBM Cloud" home page and click on Resource list



2. locate Watsonx Assistant in the resource list

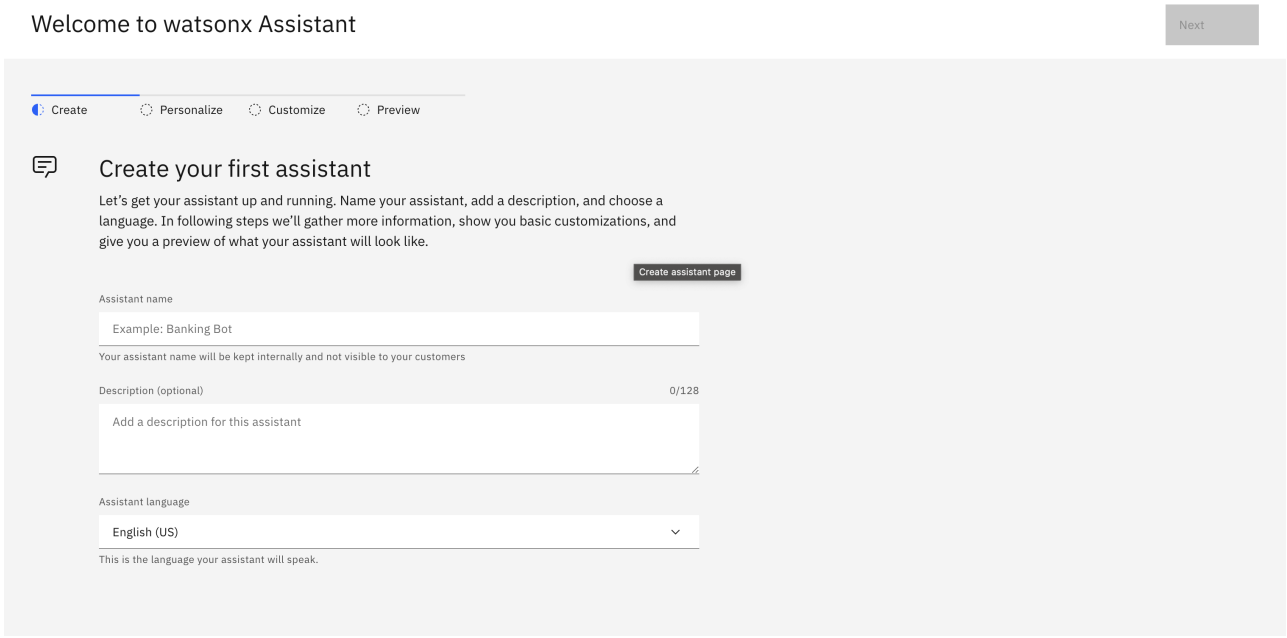


3. Launch the Watsonx Assistant

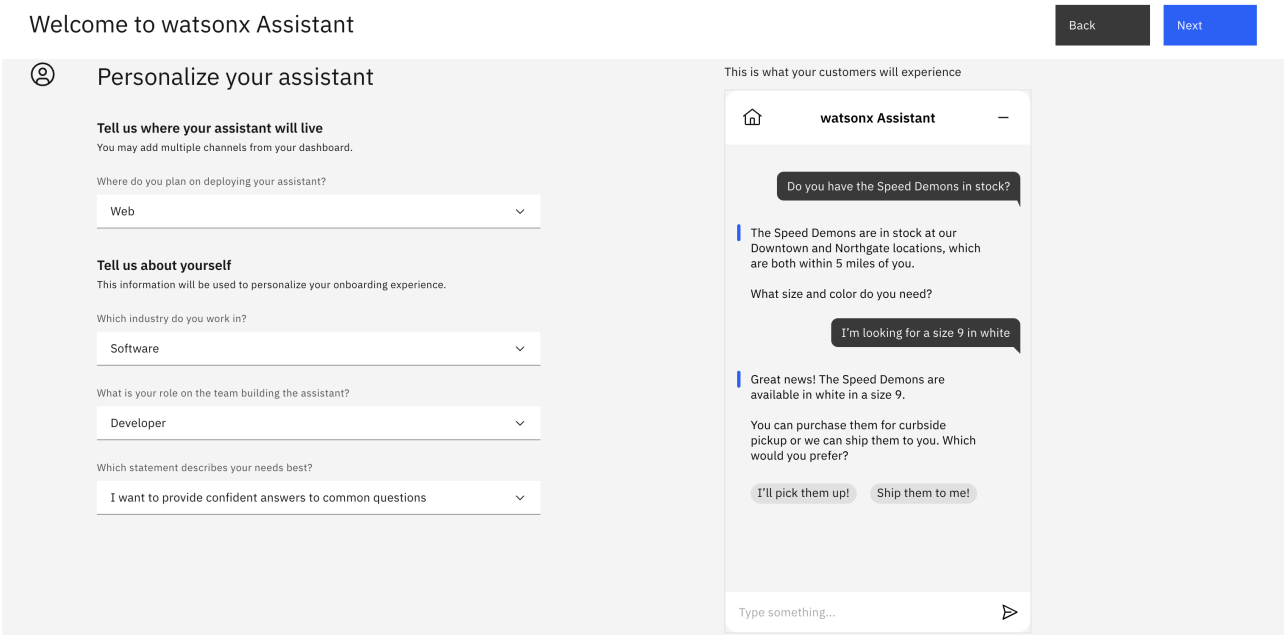


If it is the first time opening the service we need to create a first assistant. Follow step 3 - 5.

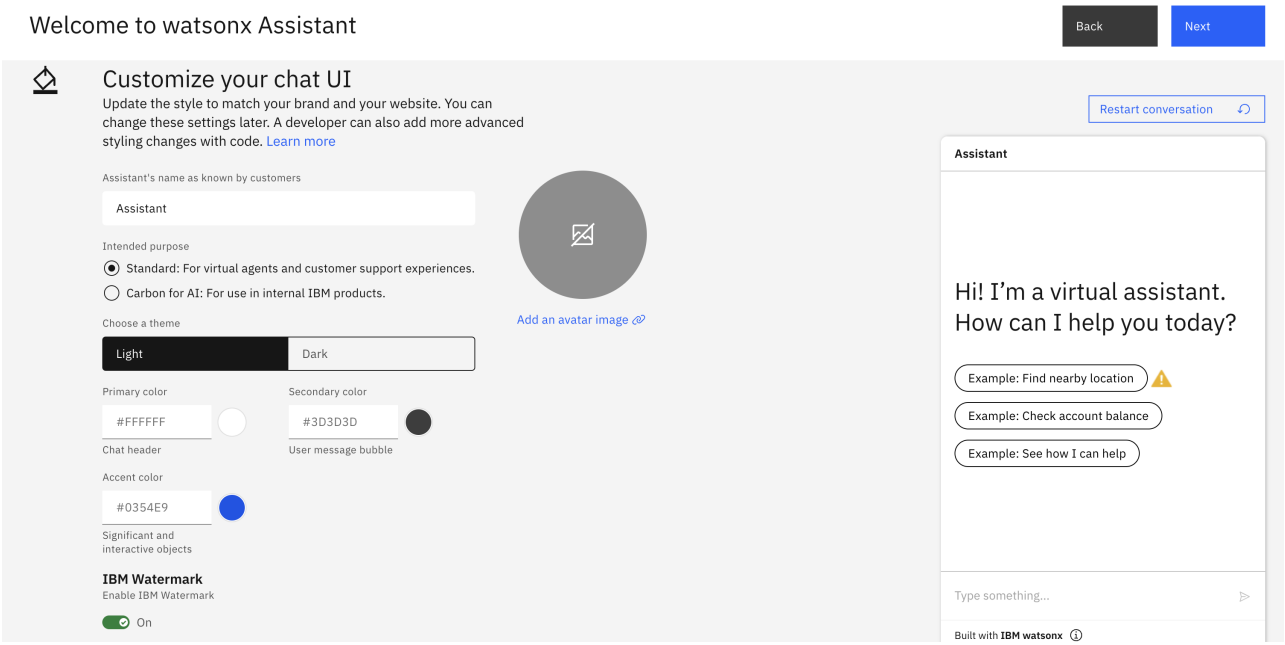
3. Give your chatbot a name with your initials and description, keep the Assistant Language in English. Click on next.



4. Choose Web in the first field. The second field can be whichever industry you prefer. The third field should be **Developer** and the last field should be **I want to provide confident answers to common questions**

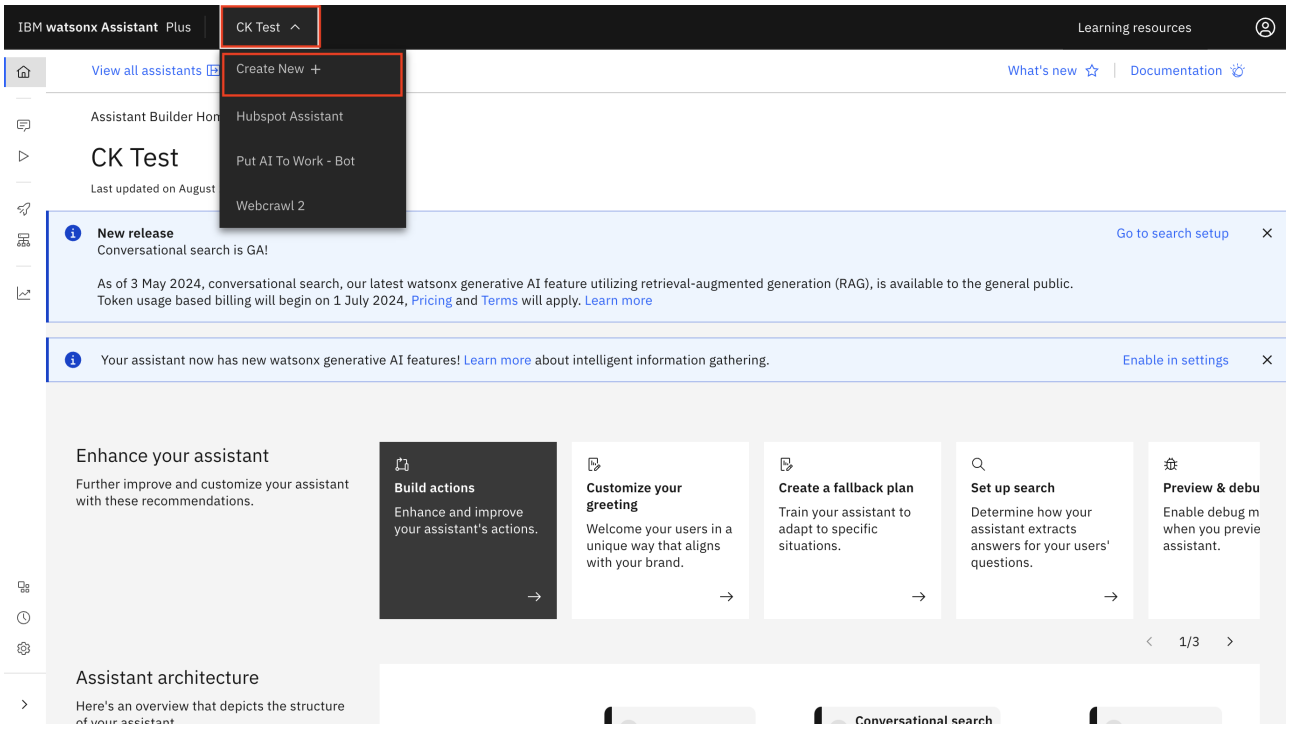


5. Customize your chatbot and press next.

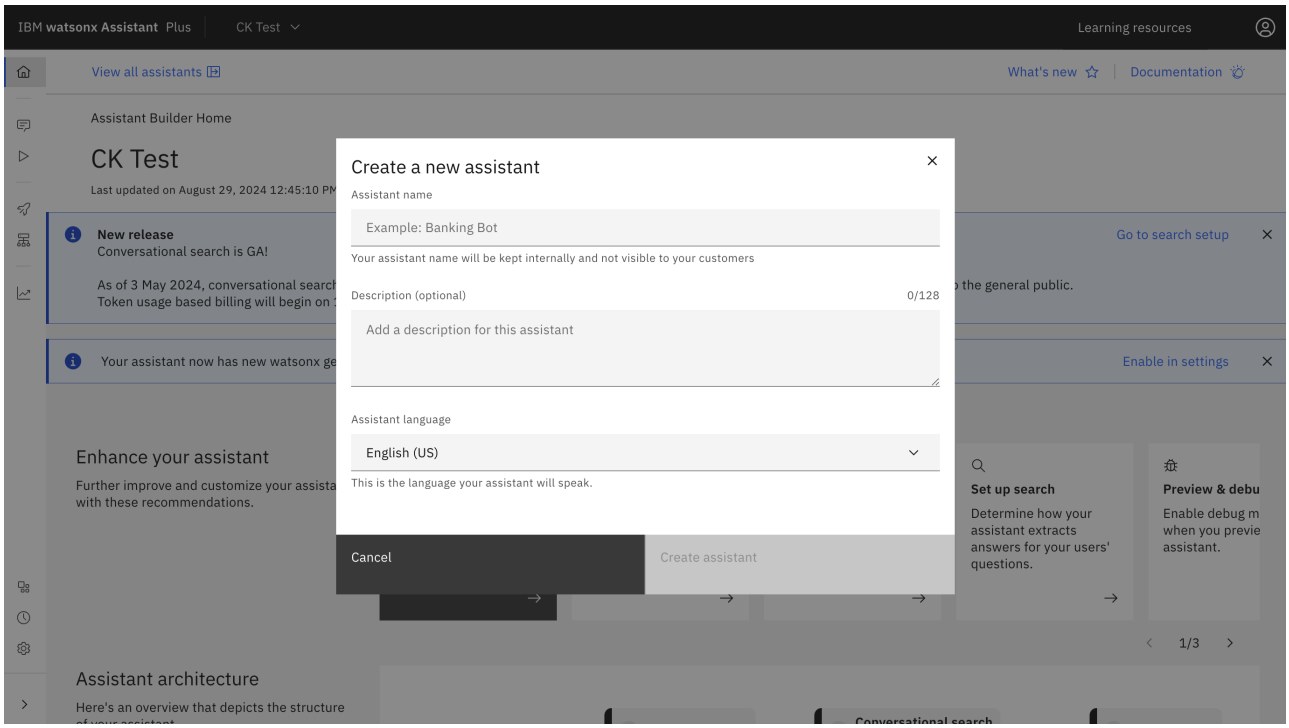


If there already exist some chatbots, follow step 6 - 8 to create your chatbot.

6. Click on the chabot name (this can be different from your view) on the navigate bar on the top to expand the folder. Click on **Create New**.



7. Give your chatbot a name with your initials and choose the languague for it.



3.2. Setting up the extension for Watsonx.ai

To build custom integration for your chatbot, we will need to create two extensions in Watsonx Assistant: 1) extension for Watsonx.ai (to connect to an LLM model); 2) extension for Watson Discovery (where our data resides). We will use the credentials gathered and saved earlier to connect Watson Discovery and Watsonx.ai services to Watsonx Assistant. We will perform these three steps in Watsonx Assistant:

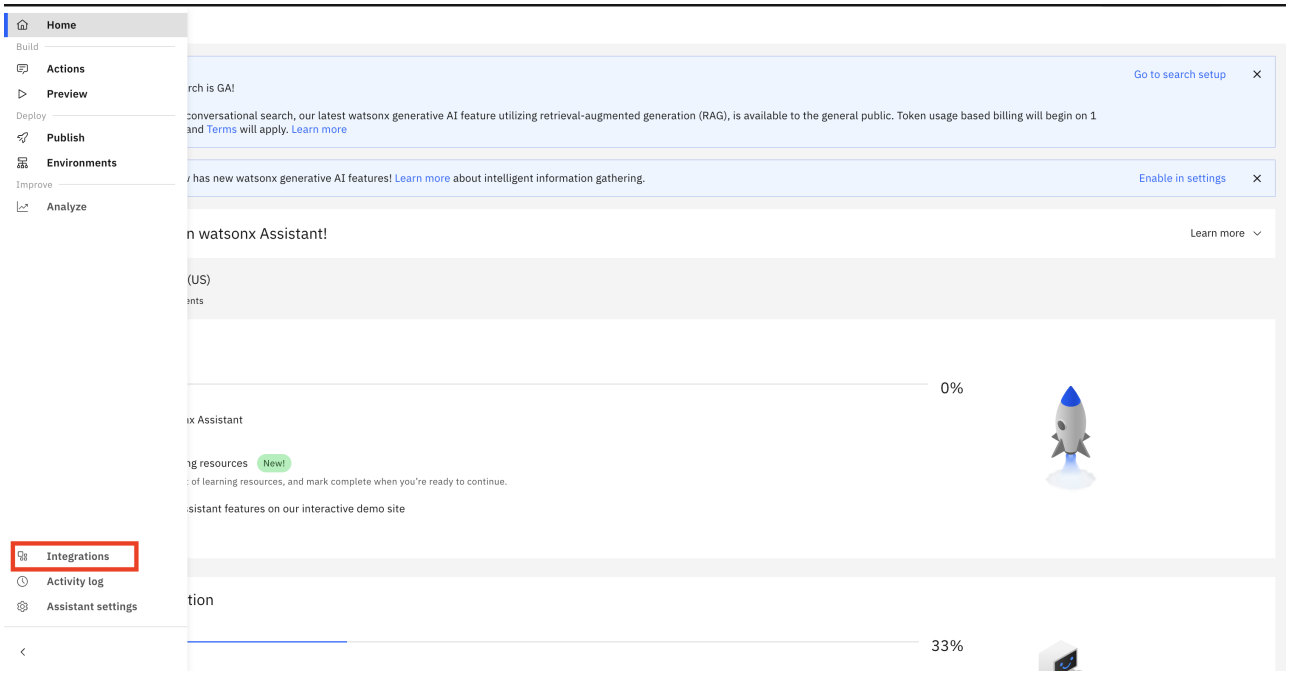
- Set up an extension to integrate Watsonx Assistant with Watsonx.ai.
- Set up an extension to integrate Watsonx Assistant with Watson Discovery.

- Create our first Watsonx Assistant actions. We will upload ready-to-use actions that created for this lab.

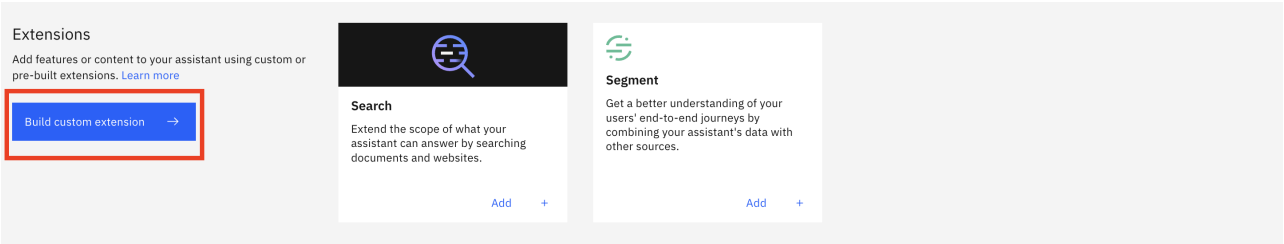
It does not matter which of the two extensions we add first. In this guide we will begin with Watsonx.ai extension.

3.2.1. Add Watsonx.ai extension

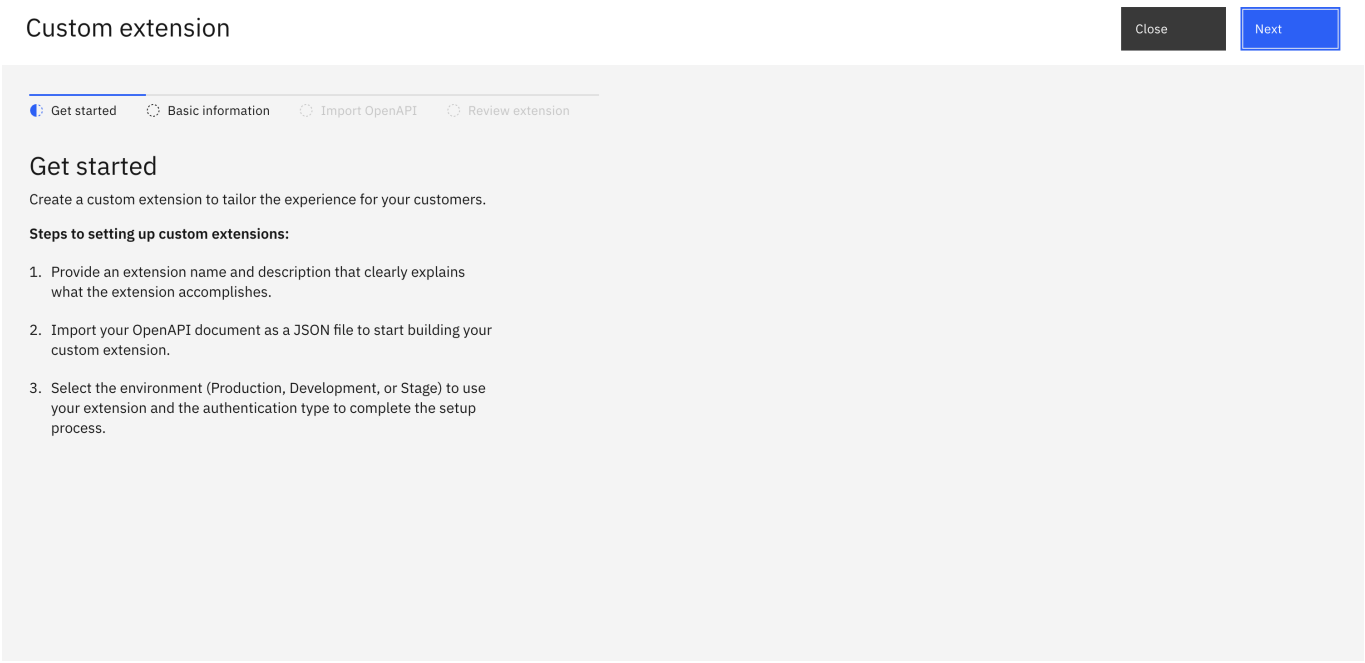
1. In the assistant menu on the left, click on **Integration**. If you can't find the menu, first click on **"IBM watsonx Assistant Plus"** on the top left corner to get back to the Home page before locating the assistant menu.



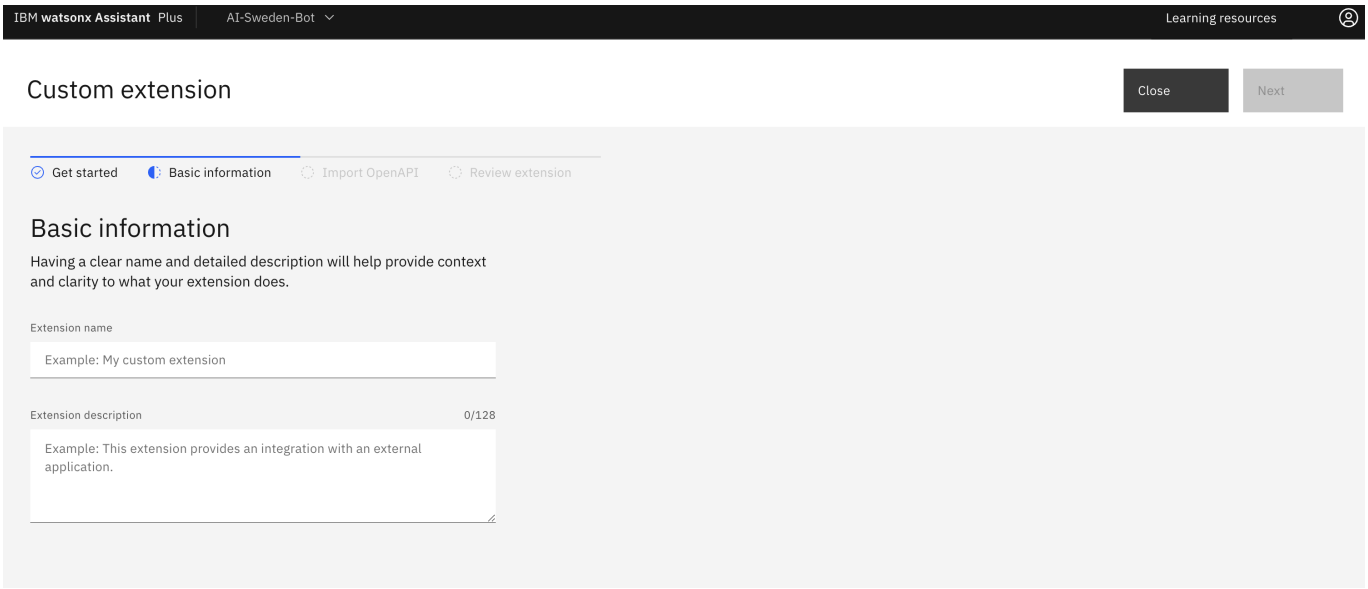
2. On next page, scroll down until you find **Extensions** and then click on the **Build custom extension** button.



3. You can start by clicking on **Next** when you see the first page.

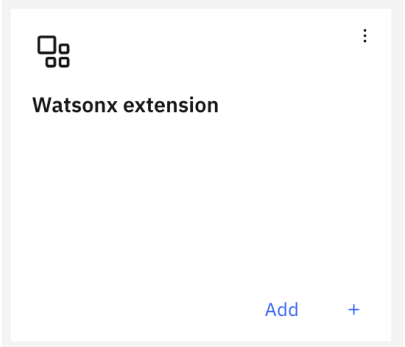


4. Give a name to your custom extension, for example *Watson Discovery extension* and optionally add a description to the extension. Click **Next** when done.

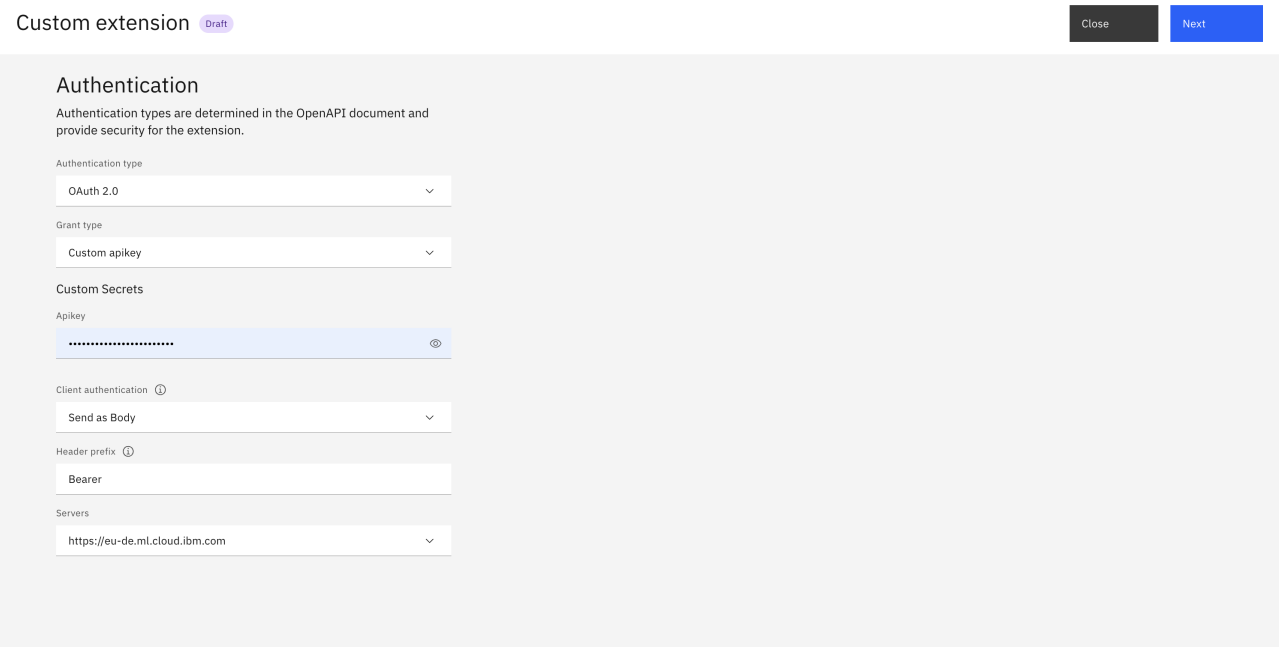


5. Upload the openAPI specification `watsonx-openapi.json` that is found here (password: Watsonx2024). Click on **Next**, and then click **Finish**.

6. Click on the **Add** button on the newly created extension and click then **Next**.



7. Change the Authentication type to *oAuth 2.0*. Add the **WX-API-key** key that you saved from 2.1. then click **Next**, then **Finish**.

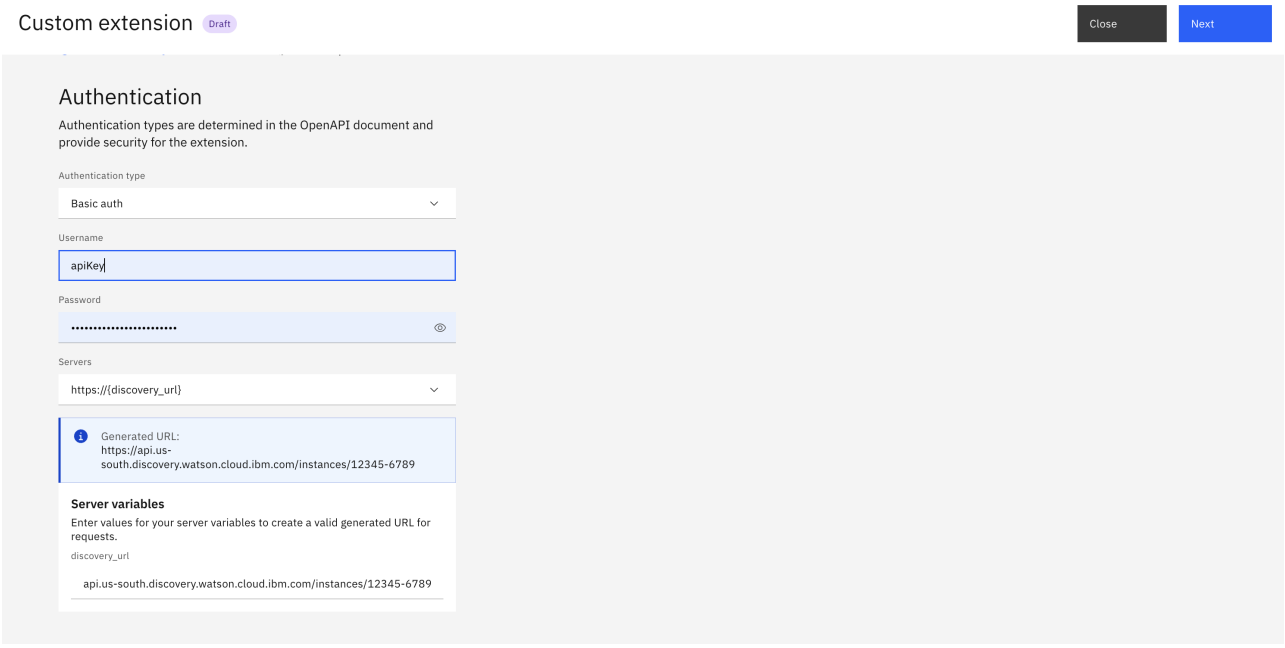


3.2.2. Add Watson Discovery extension

Next thing we will do is create a Watson Discovery extension.

1. Click on **Build custom extension.** again.
2. Name your extension, for example *Watson Discovery extension* and optionally add a description.
3. Upload the openAPI specification for the Watson Discovery extension `watsondiscovery-query-openapi.json` which you can be find here (password: Watsonx2024). Click on **Next**, then **Finish**.
4. Click on the **Add** button on the Watson Discovery extension. Click **Next**, then change the authentication type to *Basic auth*.
5. Write "**apiKey**" as the username.
6. Add the **WD API key** that you saved from section 1.1 in the password section and in the Server variables add the **WD URL** found in Watson Discoveries launching site.

   Note: Remove the **https://** part of the URL in the *Server variables*.



Once you're done, click **Next** and then **Finish**

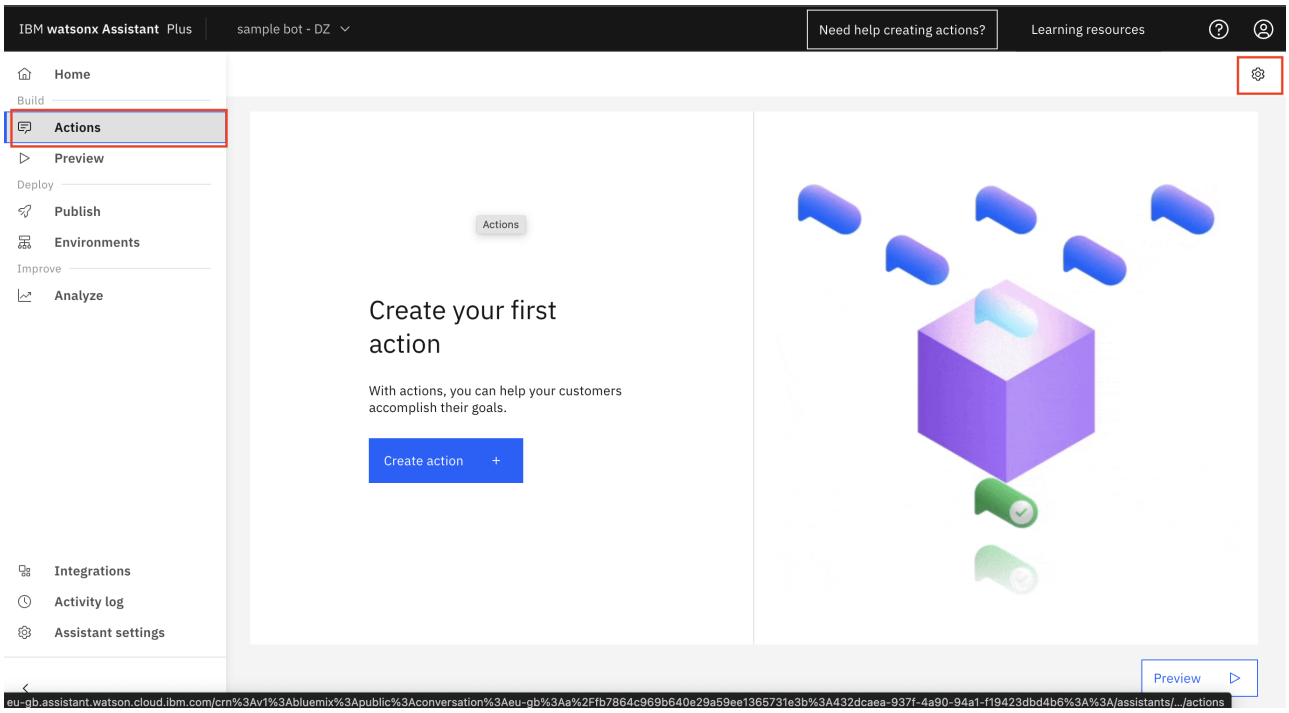
3.4. Set up actions

An action represents a discrete outcome that we want our assistant to be able to accomplish in response to a user's request. An action comprises the interaction between a customer and the assistant about a particular question or request. Actions allow us to define tasks that the assistant can perform, such as calling external services like Watson Discovery and Watsonx.

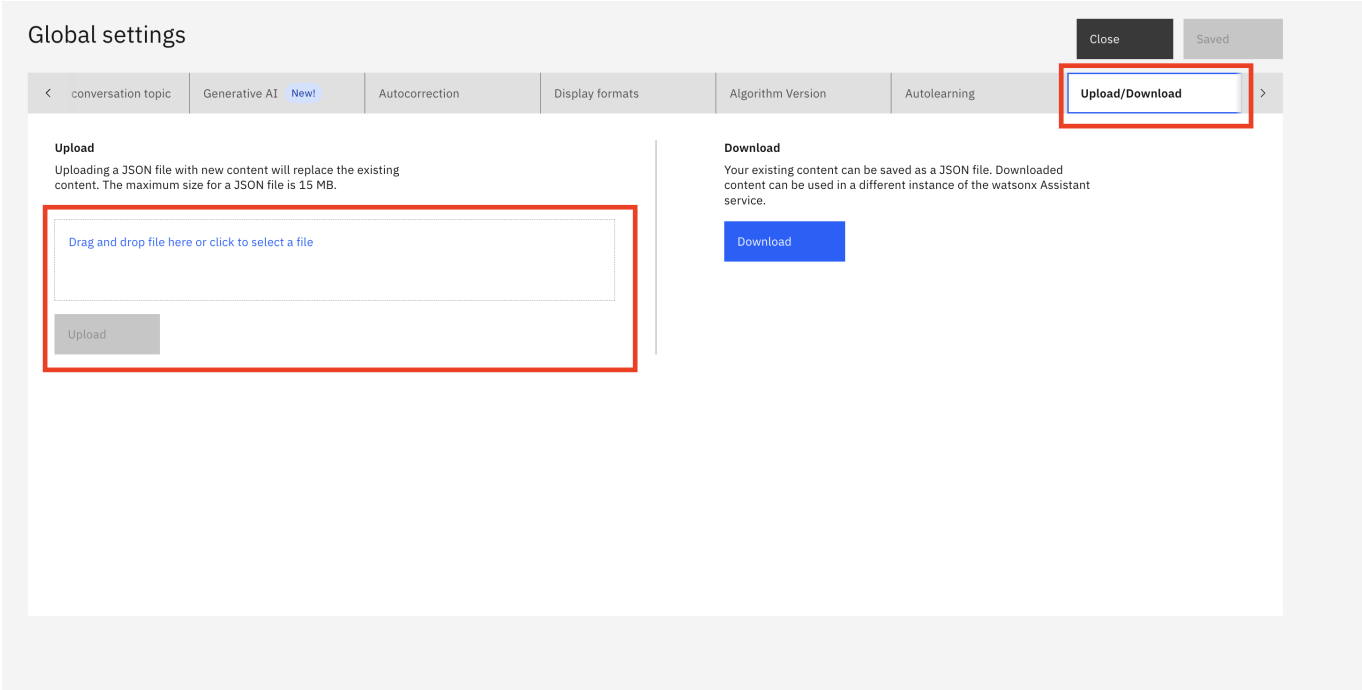
3.4.1 Set up pre-defined actions

When creating your actions in Watsonx Assistant, you can created them from scratch or import existing actions that someone already created for another assistant. The latter approach is what we call using pre-defined actions. To upload pre-defined actions, do the following:

1. Go back to the home screen of Watsonx Assistant. In the menu on the left, click on the **Actions** menu option. Click the cogwheel logo in the top right corner of the page.



2. Click on the **Upload/Download** tab as shown in image bellow.

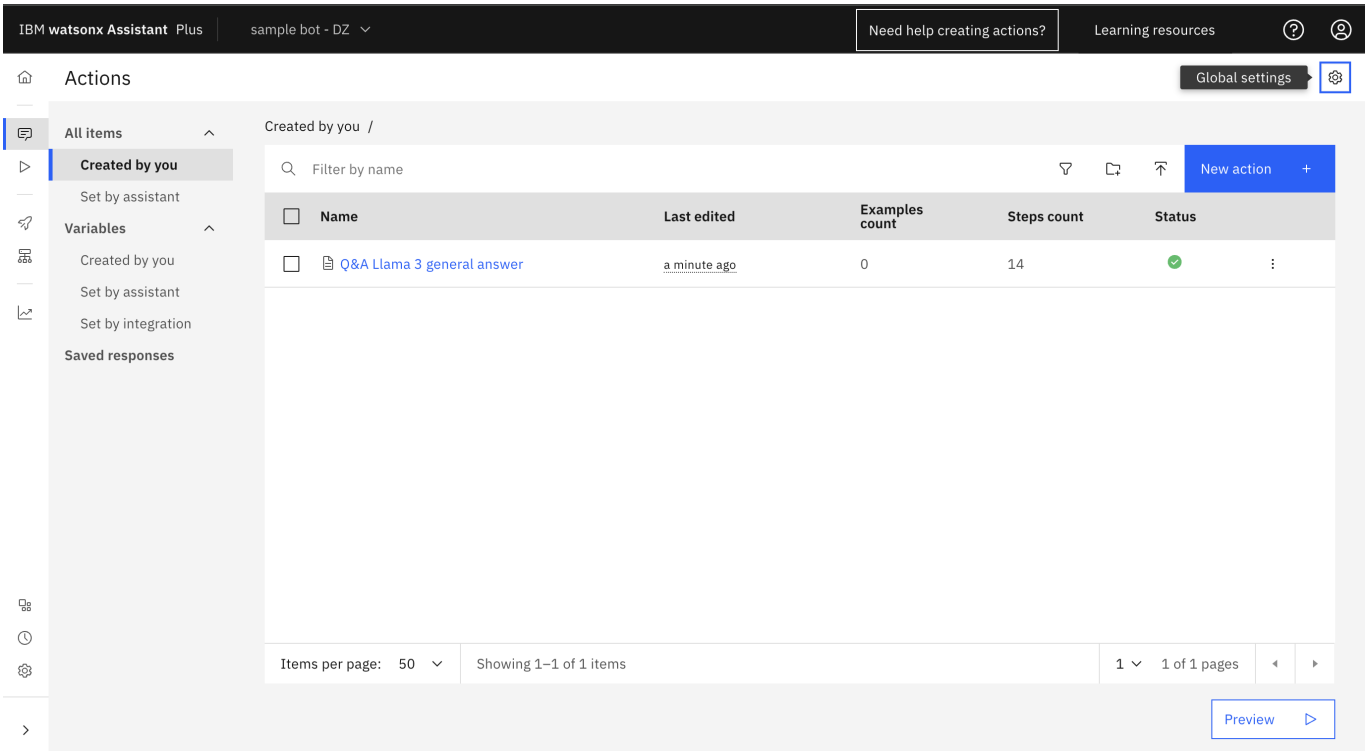


3. Click on the **Upload** button, or drag and drop `Action_AI.json` here (password: Watsonx2024) which contains downloaded actions.

4. When the action is uploaded press the **saved** button in the top right corner.

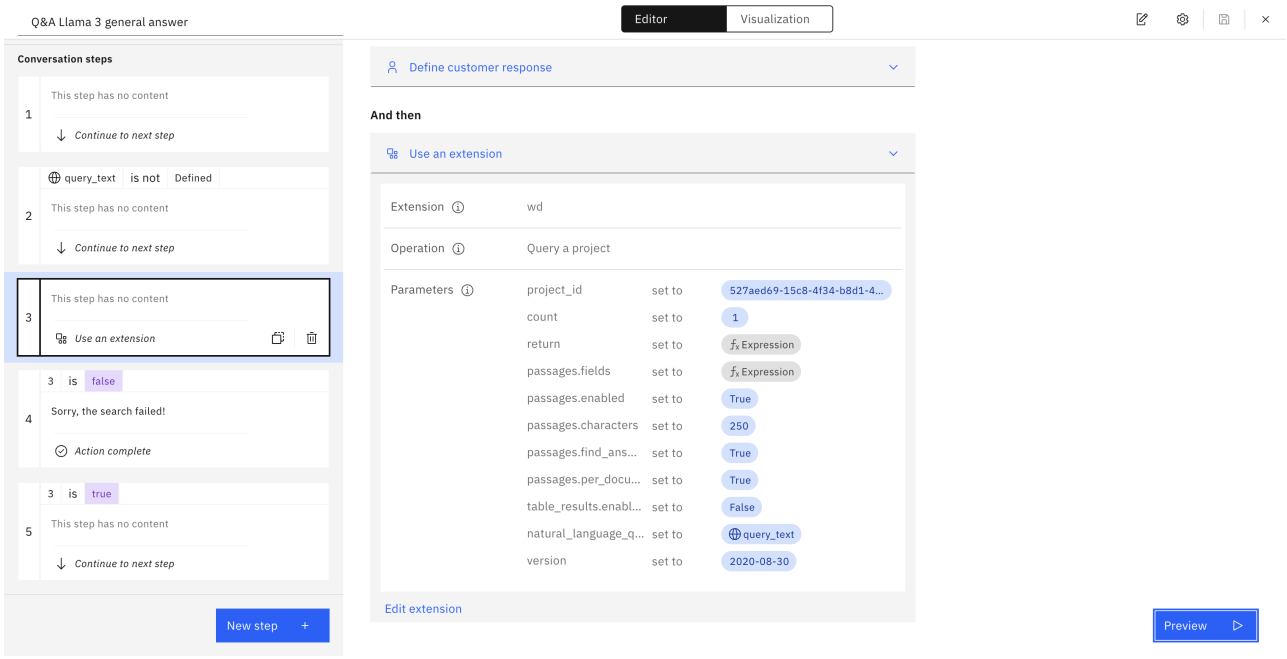
3.5. Getting the assistant to answer questions

Once you have actions in your assistant, they will look similar to the picture below.



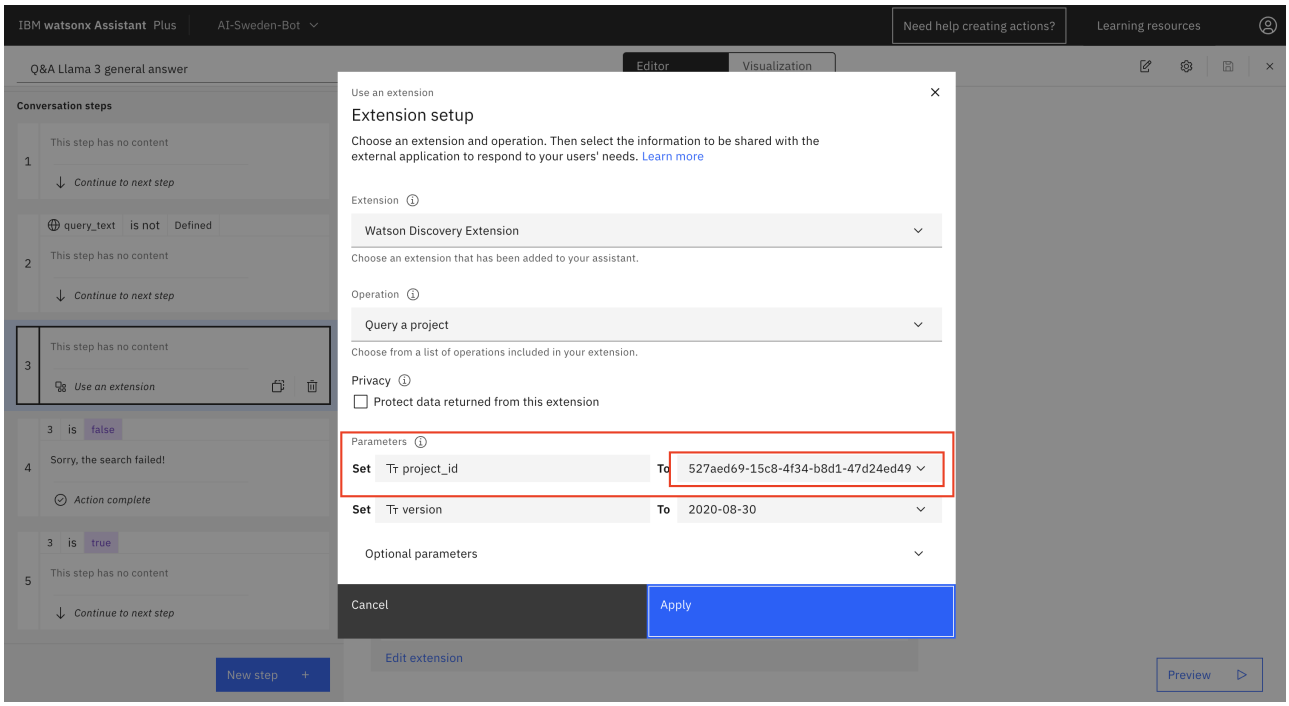
Do the following to config variables:

1. Click on the uploaded action called **Q&A Llama 3 general answer**. Then go to step number 3 as shown in the picture below.



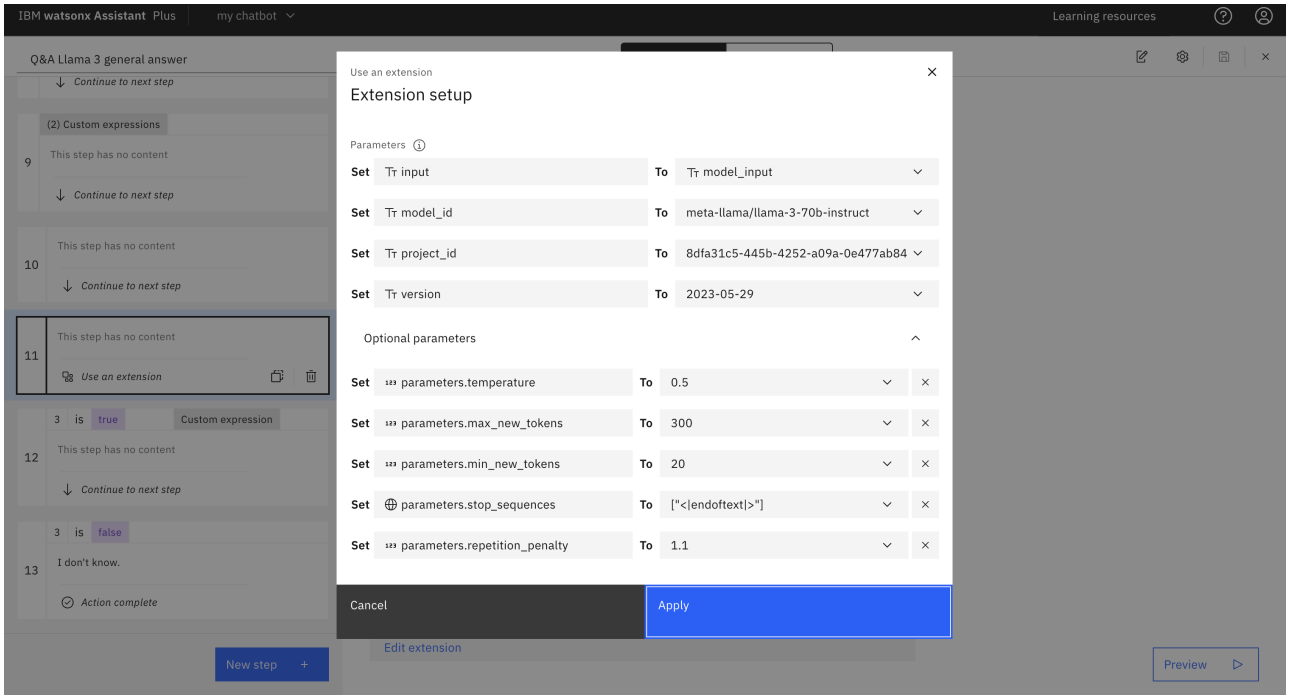
2. Click on **Edit extension** on the bottom. Go to the third step of this action, *Parameters* - this is where Watson Discovery is used.

3. Change the project ID to the same we got in section 1.3 (should be **WD-Project-ID**).



And then click on **Apply**

4. Go to the step 11 of the action - this is where we call Watsonx.ai.



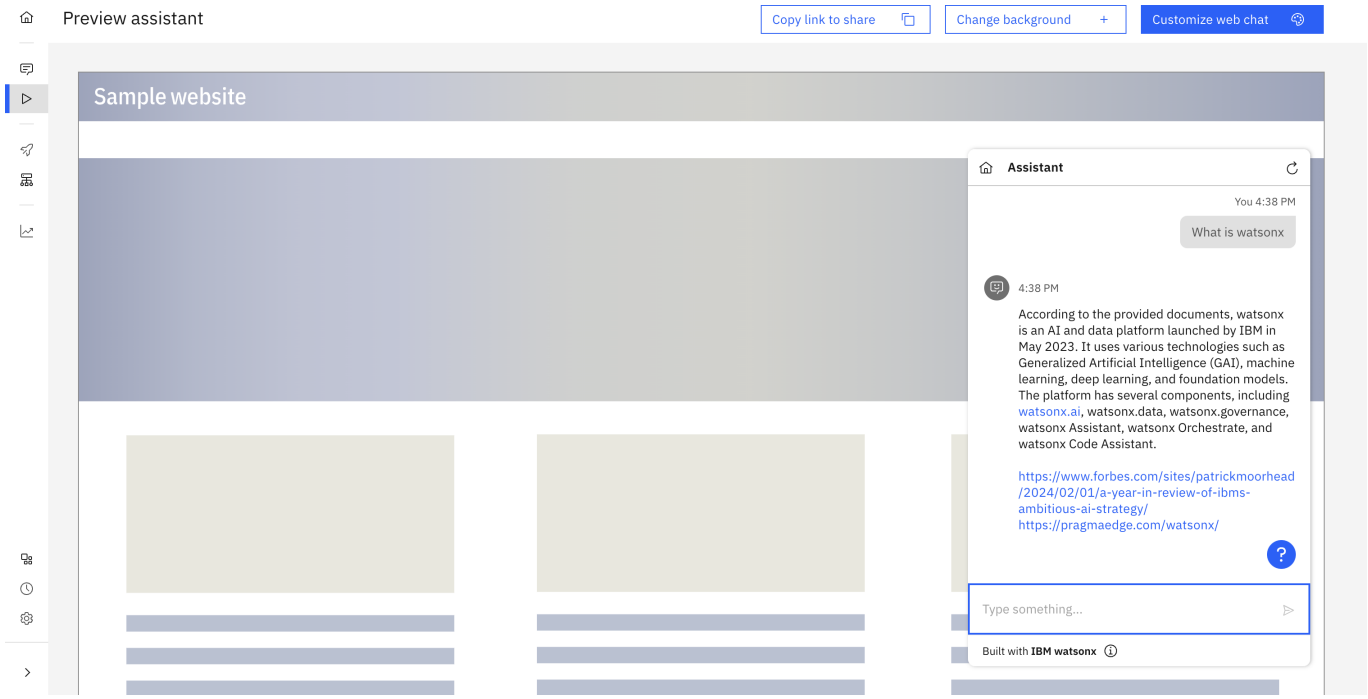
5. Click "Edit extension" and change the Watsonx.ai project ID to the project ID you got in 2.1. (named **WX-project-ID**). And then click on **Apply**.

6. Click on the save button which looks the **floppy disk icon** in the top right corner.

4. Test the chatbot

Whoho! We have successfully created a chatbot.

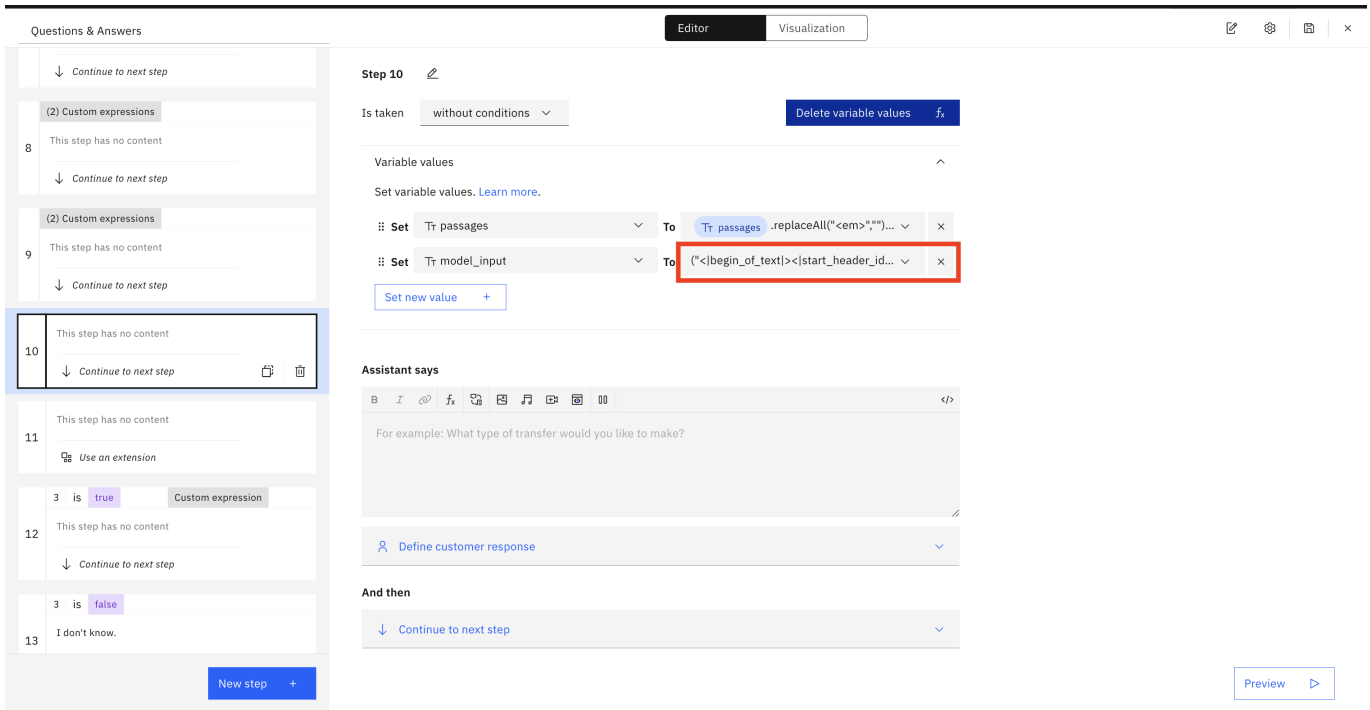
In menu on the left, click on the **Preview**, then on the right hand side in the assistant window, ask a question.



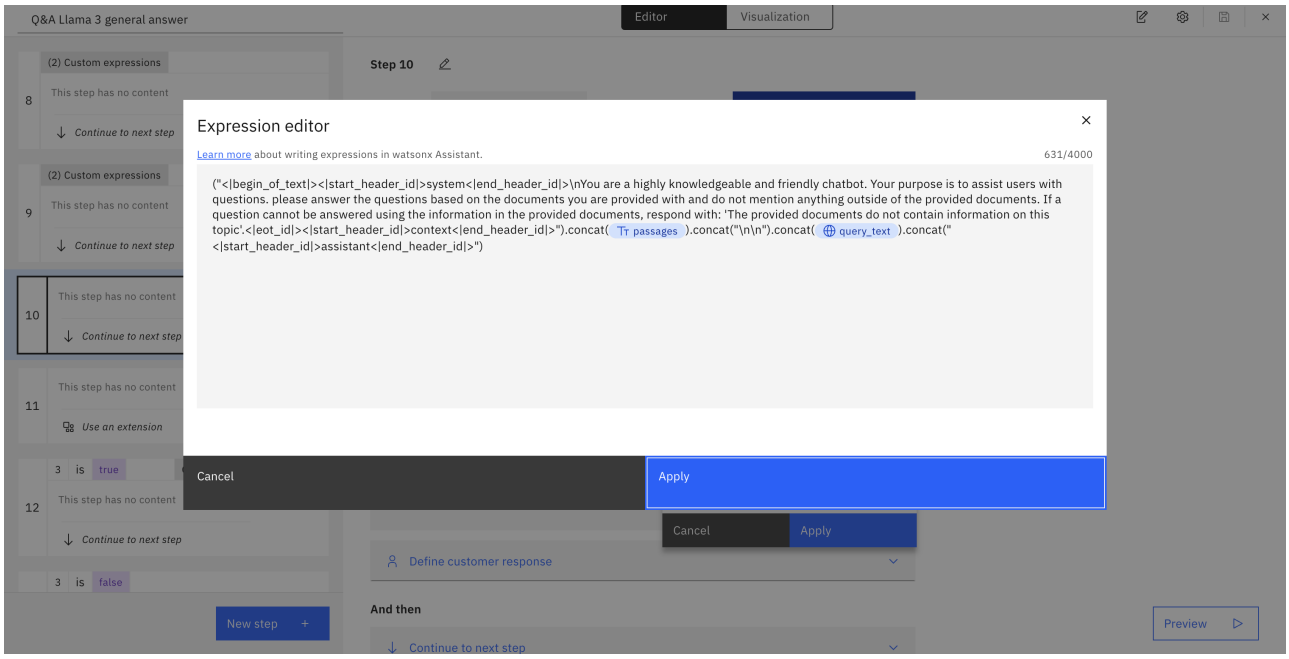
5. Optional: prompting in Watsonx Assistant

If you want to change the prompt for the model used by the chatbot do the following:

1. Go to the 10th step of the action, then click on the field marked with red square in the picture below.



2. Inside the text block is the prompt used to improve the performance of the large language model.



You may want to change the prompt for various reasons such as to get even better results, for other use cases or to support other languages.

6. Finish

Thank you for participating in this lab! We hope you enjoyed exploring **Watsonx.ai**, **Watson Assistant**, and **Watson Discovery**, and that you found the process of building a virtual assistant with Retrieval-Augmented Generation (RAG) insightful. These powerful tools are designed to help you push the boundaries of AI-driven solutions.

We encourage you to continue experimenting, innovating, and discovering new ways to integrate AI into your business.

We look forward to seeing the amazing things you will build with our products in the future. Enjoy the rest of your journey, and thank you again for your participation!