watsonx Orchestrate Lab 2:
Building a virtual assistant and setting up conversational search

Use Case

In this lab you will learn how to set up the out-of-the-box conversational search using the Retrieval-Augmented Generation (RAG) pattern by connecting the assistant to an existing knowledge base (Elasticsearch) to search for relevant FAQs and generate an answer to the user's query using an LLM in watsonx.ai. You will also learn how to configure your assistant to provide general-purpose answering using an LLM and how to set up AI-guided actions to answer questions about a specific subject using provided knowledge.

The high-level steps to accomplish this are as follows:

1. Create a new assistant builder instance
2. Configure the conversational search extension in the assistant instance
3. Test the virtual assistant's content-grounded answering by asking questions contained in the FAQ document (stored in the Elasticsearch knowledge base) as well as additional questions answered by LLM only (general-purpose answering).
4. Create and test an AI-guided action

Assumptions

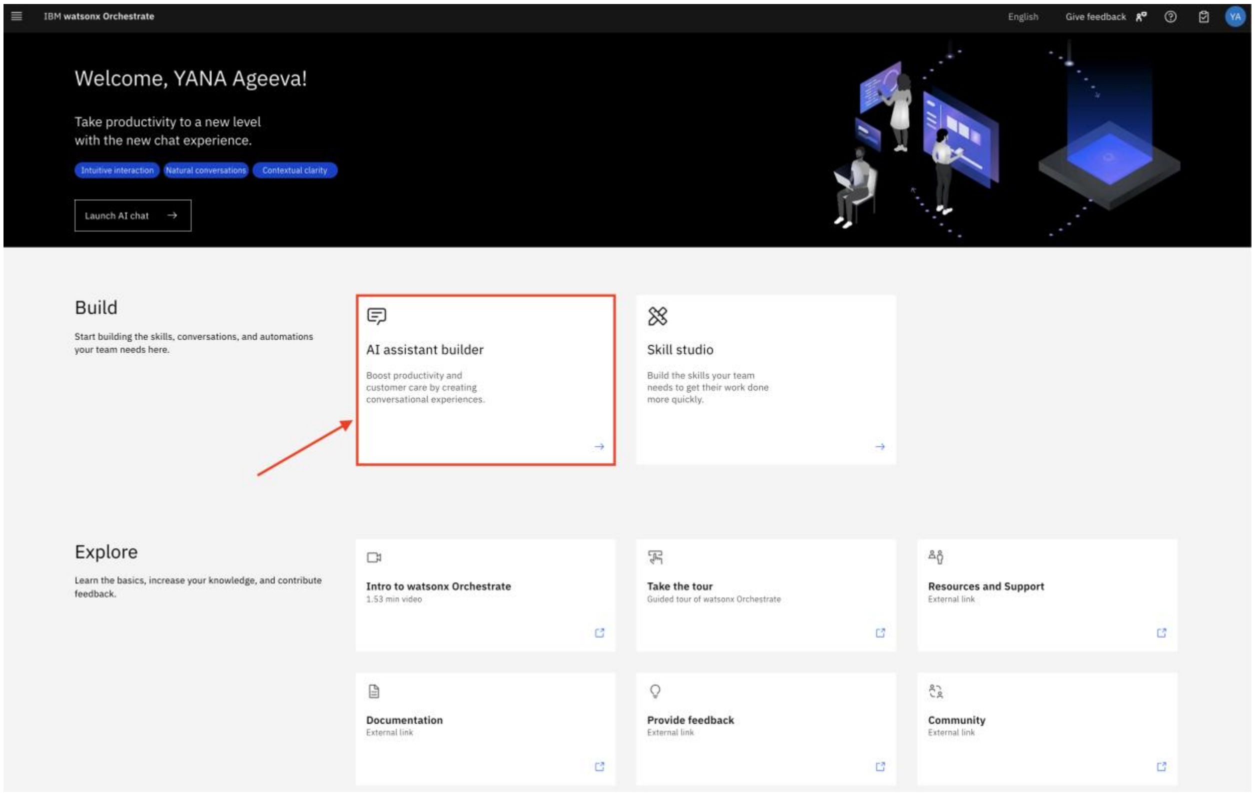
1. You have access to an IBM watsonx Orchestrate tenant (on AWS or IBM Cloud)
2. To log in, use your IBM ID and password
3. You have access to an Elasticsearch instance
4. Elasticsearch URL, port, username, and password

© 2024 IBM Corporation

Configure out-of-the-box conversational search using Elasticsearch

Create a new virtual assistant

1. Using the environment that is provided to you, log in to watsonx Orchestrate.
2. Click on the Assistants tile to open the Assistant Builder:



Welcome, YANA Ageeva!

Take productivity to a new level with the new chat experience.

Build

Start building the skills, conversations, and automations your team needs here.

AI assistant builder

Boost productivity and customer care by creating conversational experiences.

Skill studio

Build the skills your team needs to get their work done more quickly.

Explore

Learn the basics, increase your knowledge, and contribute feedback.

Intro to watsonx Orchestrate
1.53 min video

Take the tour
Guided tour of watsonx Orchestrate

Resources and Support
External link

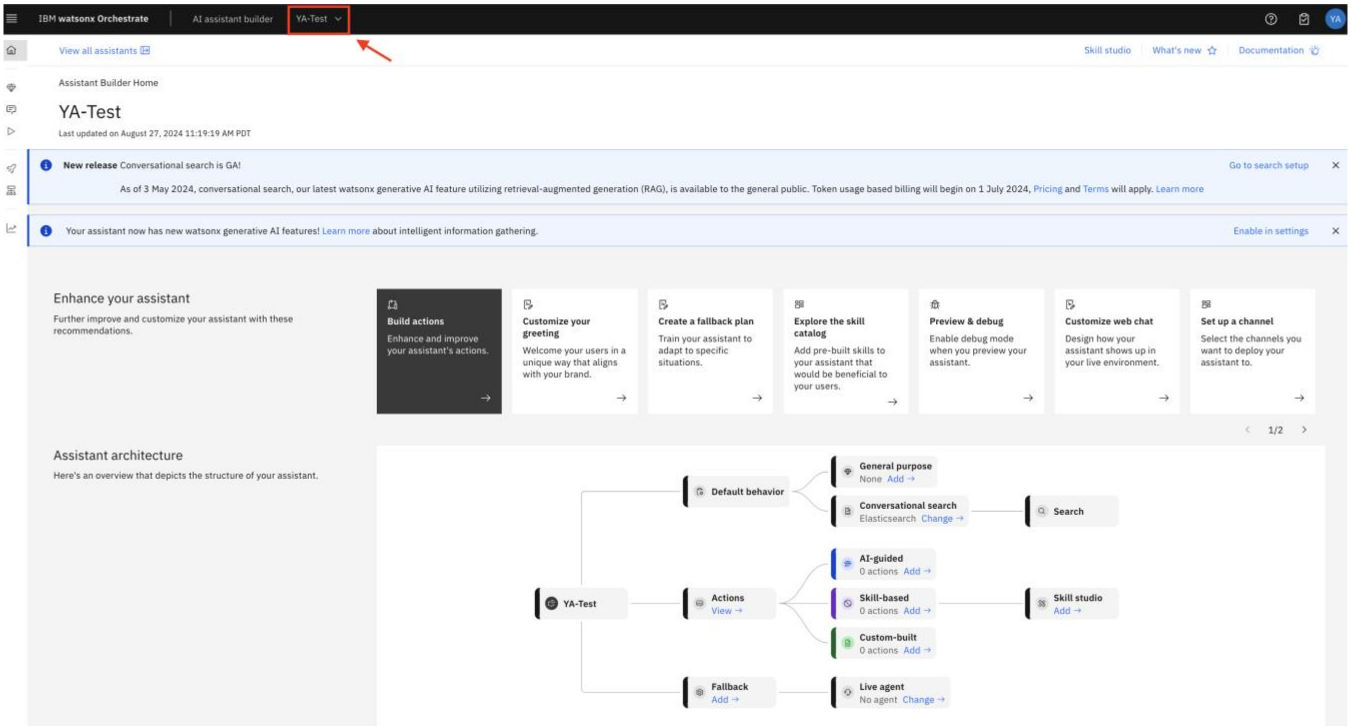
Documentation
External link

Provide feedback
External link

Community
External link

3. You should see the following screen for one of the existing assistant instances. Click on the dropdown menu at the top to see existing instances:

© 2024 IBM Corporation



New release Conversational search is GA!

As of 3 May 2024, conversational search, our latest watsonx generative AI feature utilizing retrieval-augmented generation (RAG), is available to the general public. Token usage based billing will begin on 1 July 2024, Pricing and Terms will apply. Learn more

Your assistant now has new watsonx generative AI features! Learn more about intelligent information gathering.

Further improve and customize your assistant with these recommendations.

Enhance and improve your assistant's actions.

Welcome your users in a unique way that aligns with your brand.

Train your assistant to adapt to specific situations.

Add pre-built skills to your assistant that would be beneficial to your users.

Enable debug mode when you preview your assistant.

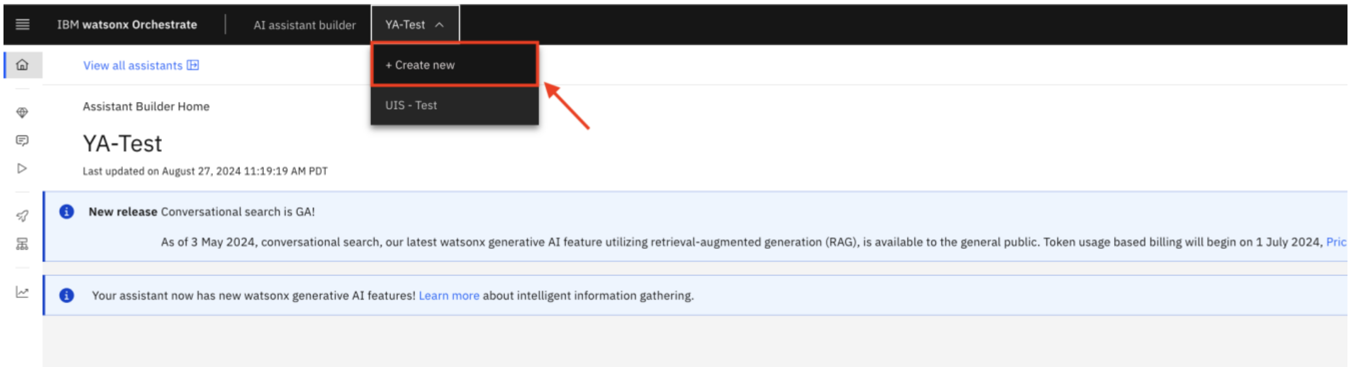
Design how your assistant shows up in your live environment.

Select the channels you want to deploy your assistant to.

Here's an overview that depicts the structure of your assistant.

Note: *if there are no existing instances, you will have to create the first one (answer the questions in the wizard and use default settings when possible.*

4. Create a new assistant instance



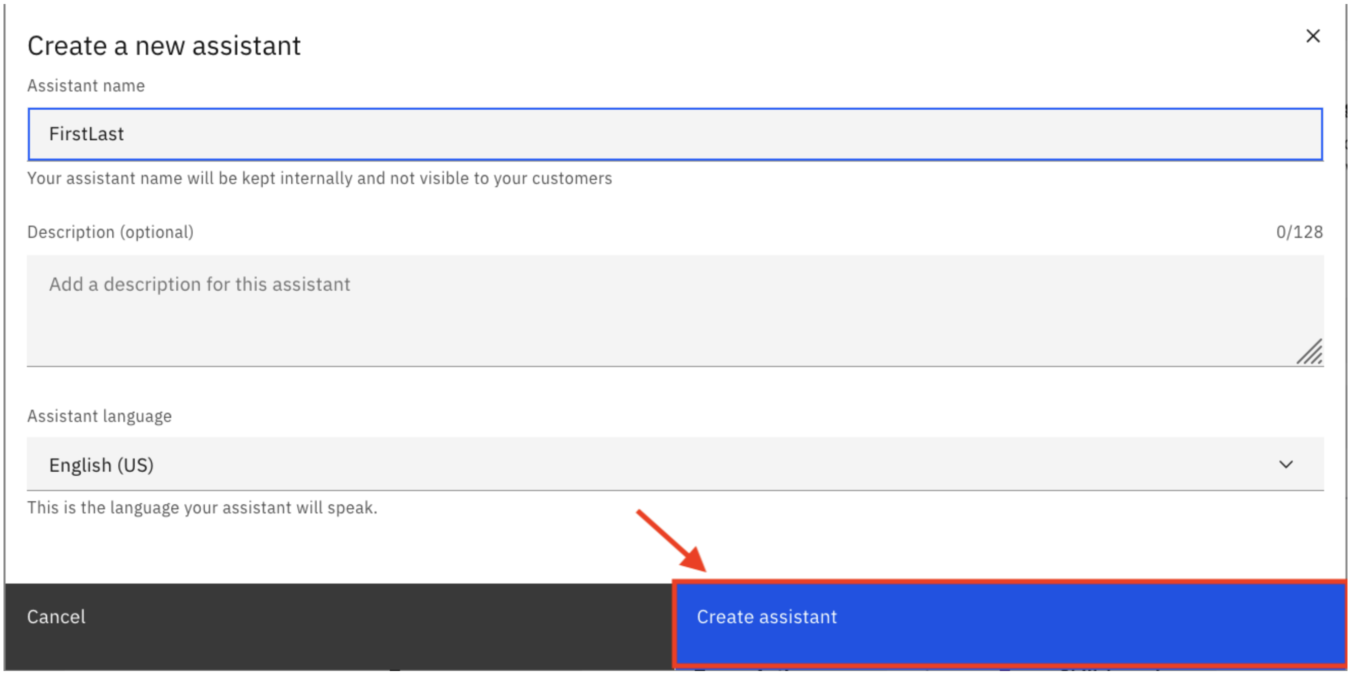
New release Conversational search is GA!

As of 3 May 2024, conversational search, our latest watsonx generative AI feature utilizing retrieval-augmented generation (RAG), is available to the general public. Token usage based billing will begin on 1 July 2024, Pric

Your assistant now has new watsonx generative AI features! Learn more about intelligent information gathering.

5. Name your assistant as FirstLast - Assistant (where First is your first name, Last is your last name) and click **Create assistant**.

© 2024 IBM Corporation



Create a new assistant

Assistant name: FirstLast

Your assistant name will be kept internally and not visible to your customers

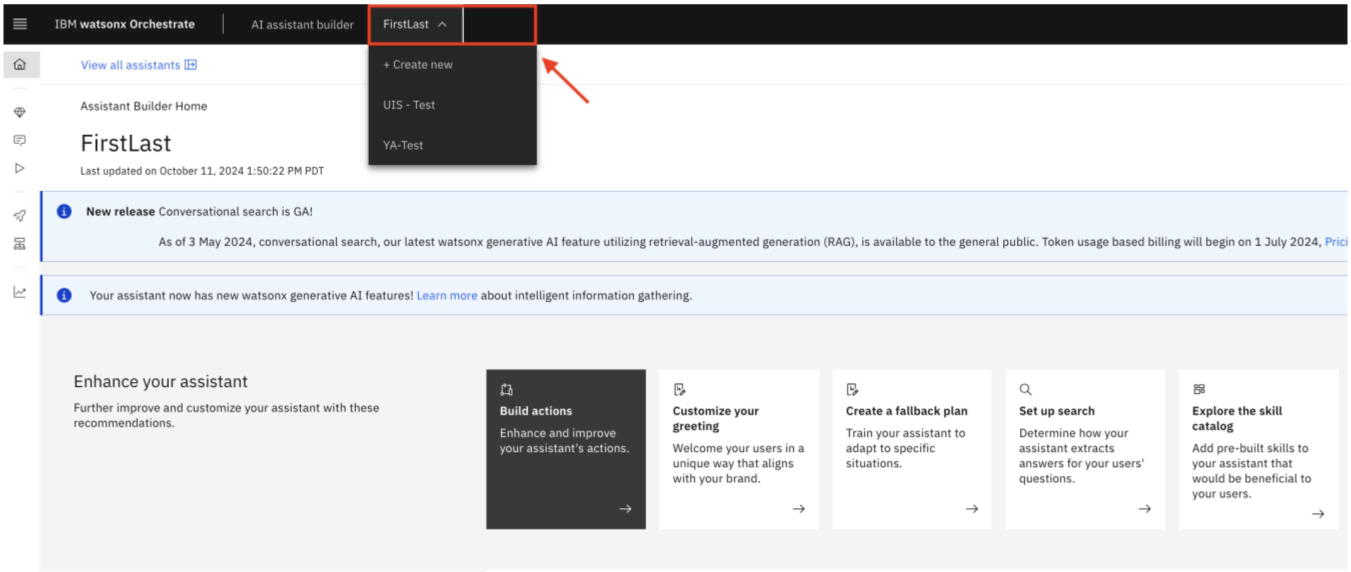
Description (optional) 0/128

Add a description for this assistant

Assistant language: English (US)

This is the language your assistant will speak.

6. Your assistant has been created and you can now see it in the dropdown list.



FirstLast

Last updated on October 11, 2024 1:50:22 PM PDT

New release Conversational search is GA!

As of 3 May 2024, conversational search, our latest watsonx generative AI feature utilizing retrieval-augmented generation (RAG), is available to the general public. Token usage based billing will begin on 1 July 2024, Prici

Your assistant now has new watsonx generative AI features! Learn more about intelligent information gathering.

Enhance your assistant

Further improve and customize your assistant with these recommendations.

Build actions Enhance and improve your assistant's actions.

Customize your greeting Welcome your users in a unique way that aligns with your brand.

Create a fallback plan Train your assistant to adapt to specific situations.

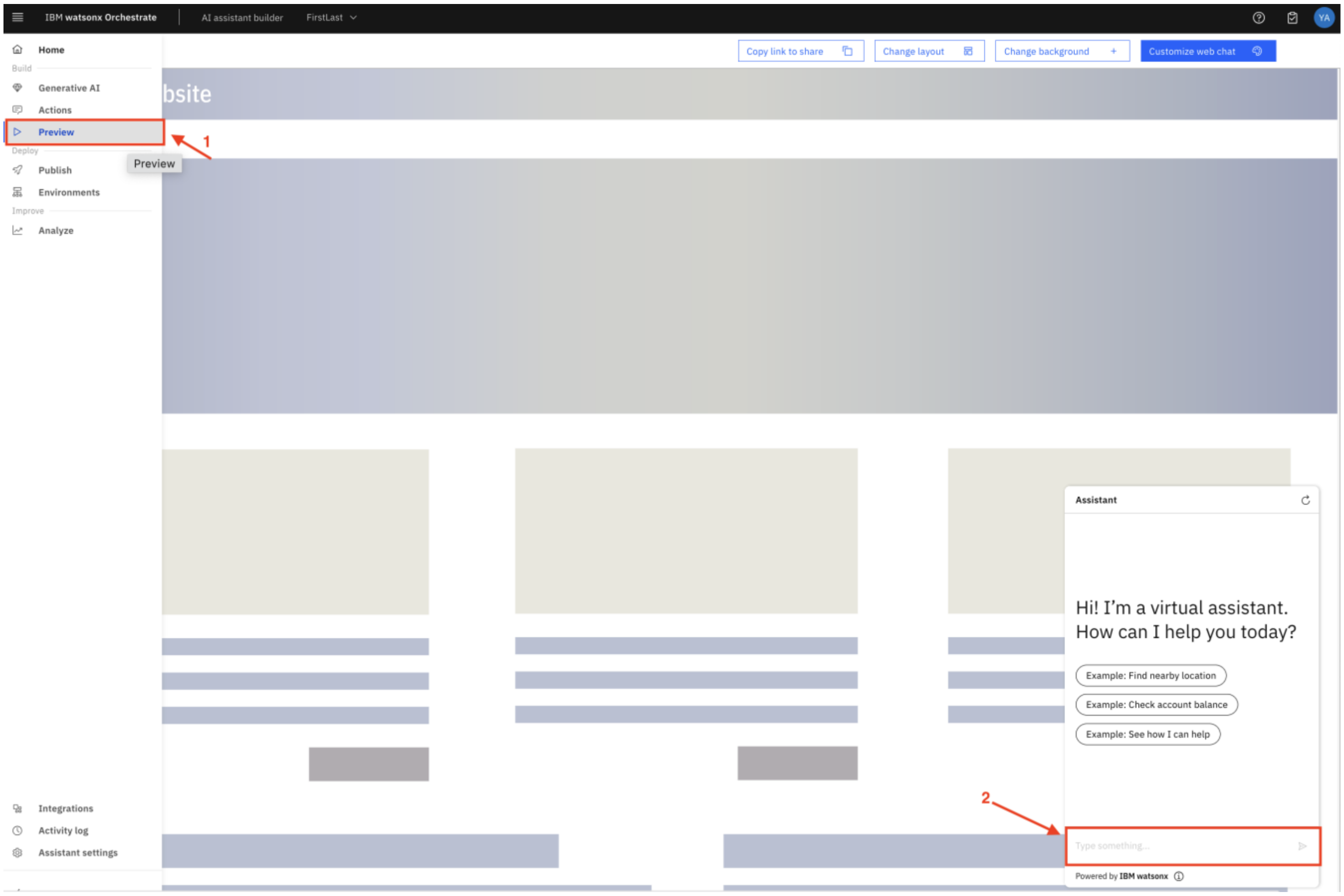
Set up search Determine how your assistant extracts answers for your users' questions.

Explore the skill catalog Add pre-built skills to your assistant that would be beneficial to your users.

Congratulations! You have successfully created your first assistant. It is connected by default to the IBM Granite LLM for general purpose answering, and you can test it by asking questions in the preview chat window.

© 2024 IBM Corporation

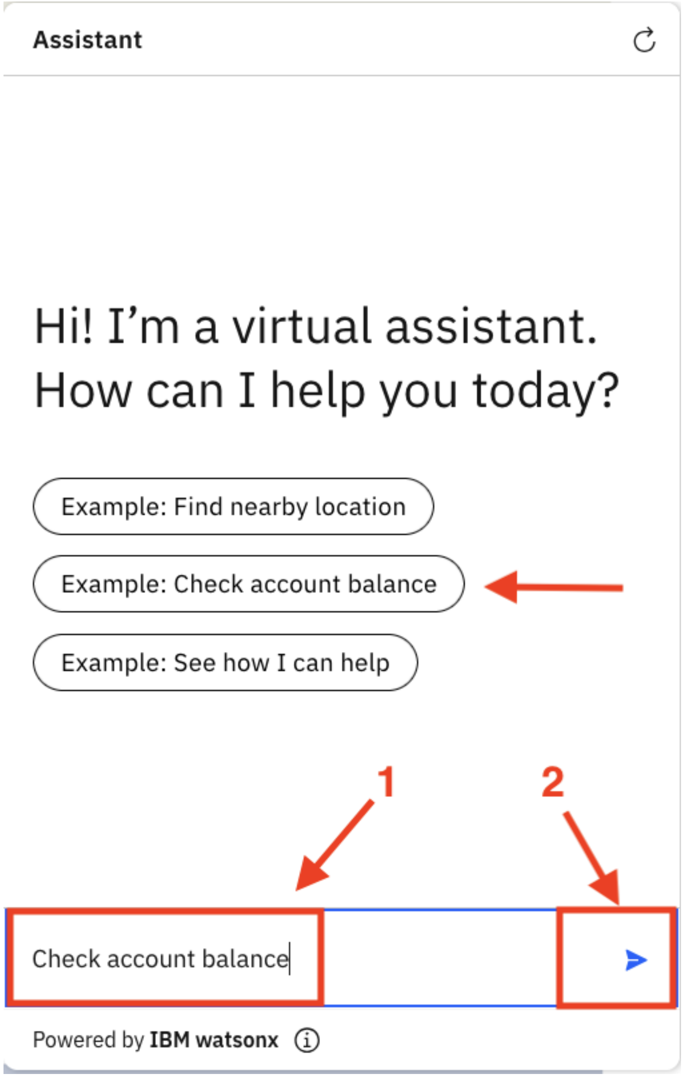
7. Click on **Preview** and move your cursor into the text area of the assistant that pops up:



Hi! I'm a virtual assistant. How can I help you today?

8. Enter any query e.g. "Check account balance" (or you can also click on one of the default buttons in the chat window)

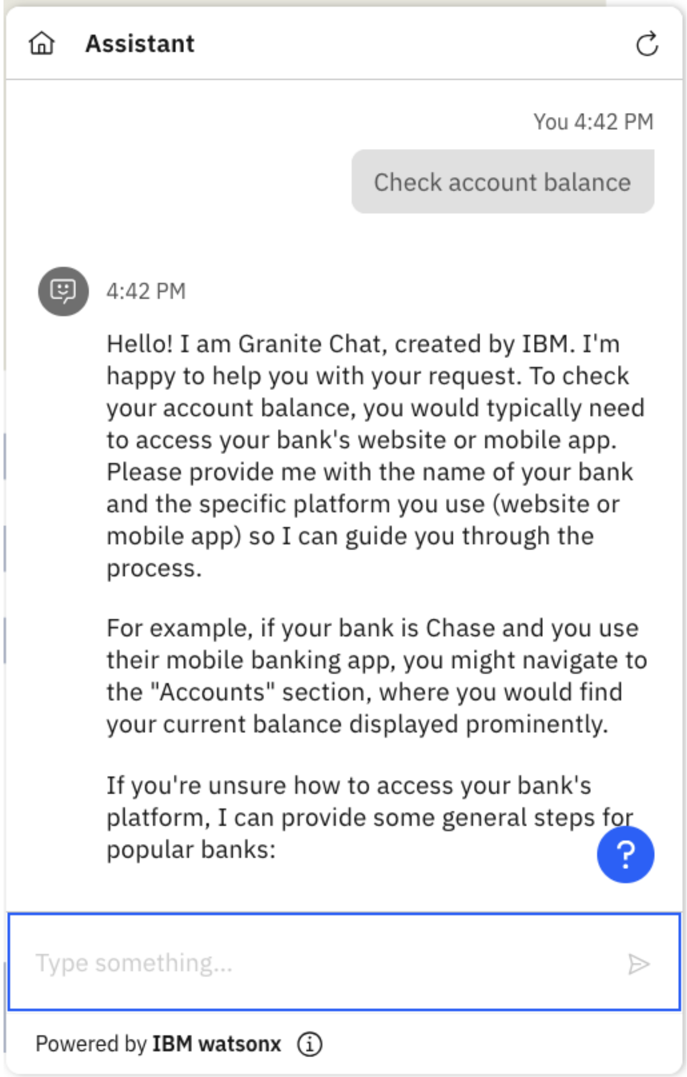
© 2024 IBM Corporation


Assistant
Hi! I'm a virtual assistant.
How can I help you today?
Example: Find nearby location
Example: Check account balance
Example: See how I can help
1
2
Powered by IBM watsonx

Check account balance

9. The answer is provided by the Granite LLM using general purpose answering:

© 2024 IBM Corporation



You 4:42 PM

Check account balance

4:42 PM

Hello! I am Granite Chat, created by IBM. I'm happy to help you with your request. To check your account balance, you would typically need to access your bank's website or mobile app. Please provide me with the name of your bank and the specific platform you use (website or mobile app) so I can guide you through the process.

For example, if your bank is Chase and you use their mobile banking app, you might navigate to the "Accounts" section, where you would find your current balance displayed prominently.

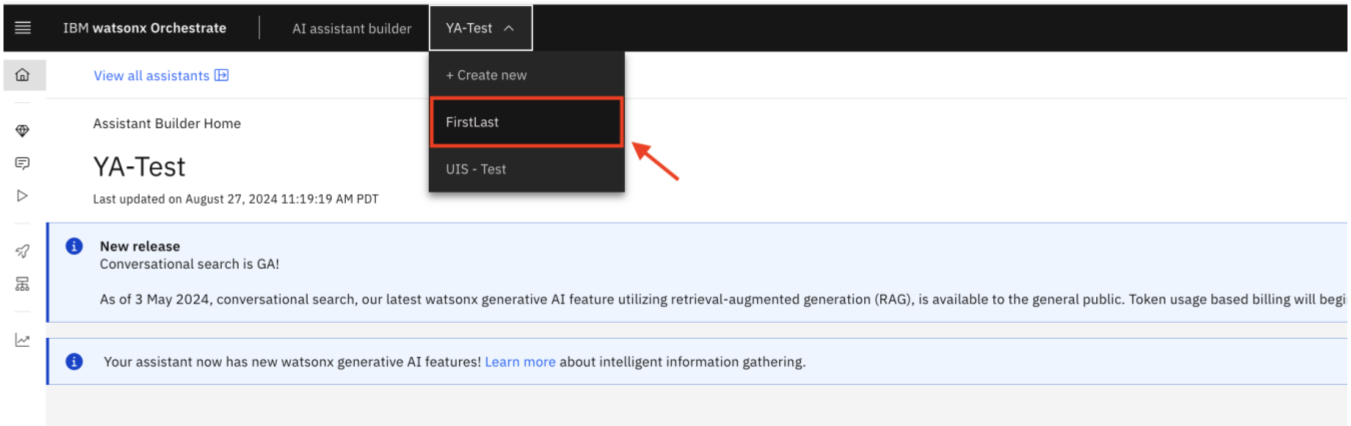
If you're unsure how to access your bank's platform, I can provide some general steps for popular banks:

As you can see, this is a very generic answer because the assistant doesn't know anything about you or which bank you have an account at. In the following steps we will set up the assistant and teach it to answer questions intelligently, based on the provided knowledge base and implemented actions.

Set up conversational search in your assistant builder instance

1. Select your assistant instance if it isn't selected already

© 2024 IBM Corporation



YA-Test

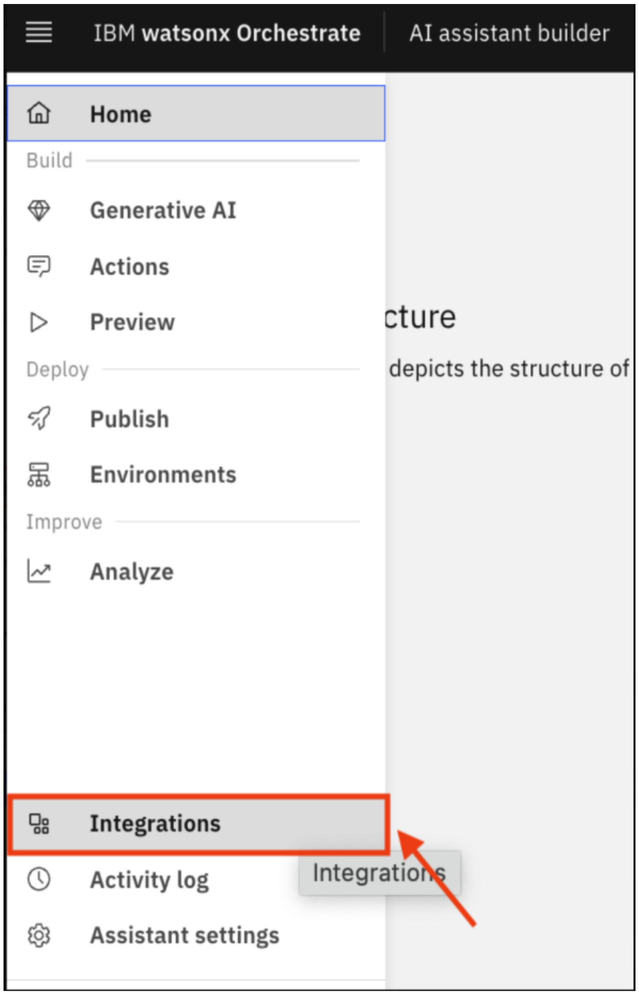
Last updated on August 27, 2024 11:19:19 AM PDT

New release
Conversational search is GA!

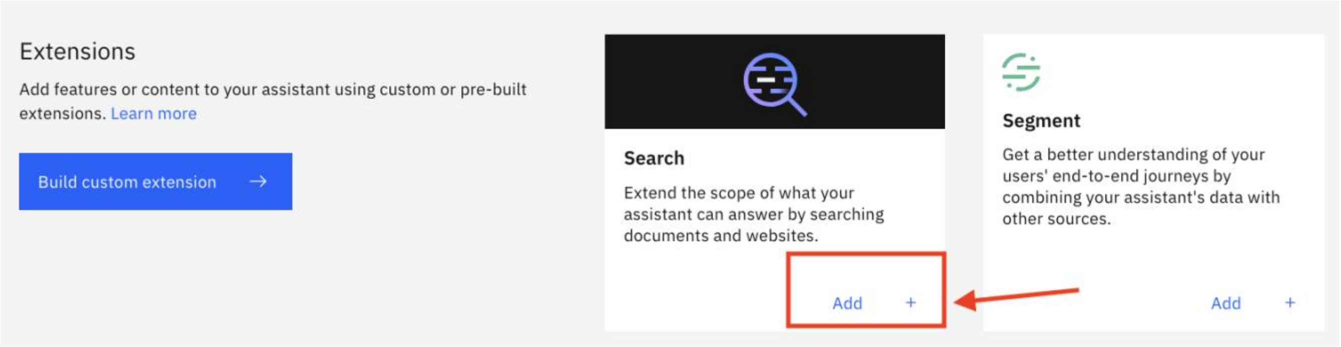
As of 3 May 2024, conversational search, our latest watsonx generative AI feature utilizing retrieval-augmented generation (RAG), is available to the general public. Token usage based billing will begi

Your assistant now has new watsonx generative AI features! Learn more about intelligent information gathering.

2. Open the **Integrations** tab



3. Scroll down to **Extensions** and add **Search**



Extensions

Add features or content to your assistant using custom or pre-built extensions. Learn more

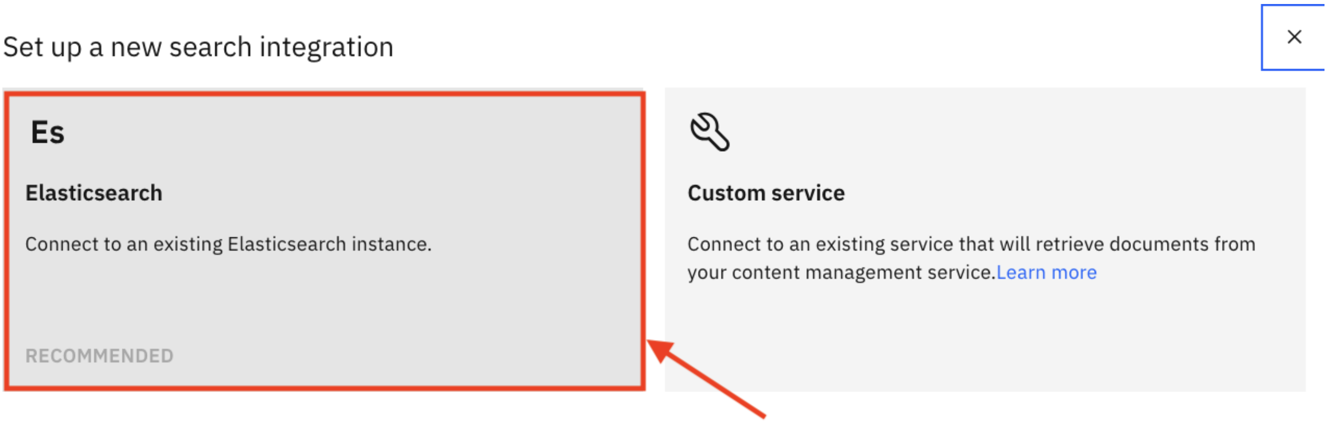
Search

Extend the scope of what your assistant can answer by searching documents and websites.

Segment

Get a better understanding of your users' end-to-end journeys by combining your assistant's data with other sources.

4. Select **Elasticsearch**



Set up a new search integration

Elasticsearch

Connect to an existing Elasticsearch instance.

RECOMMENDED

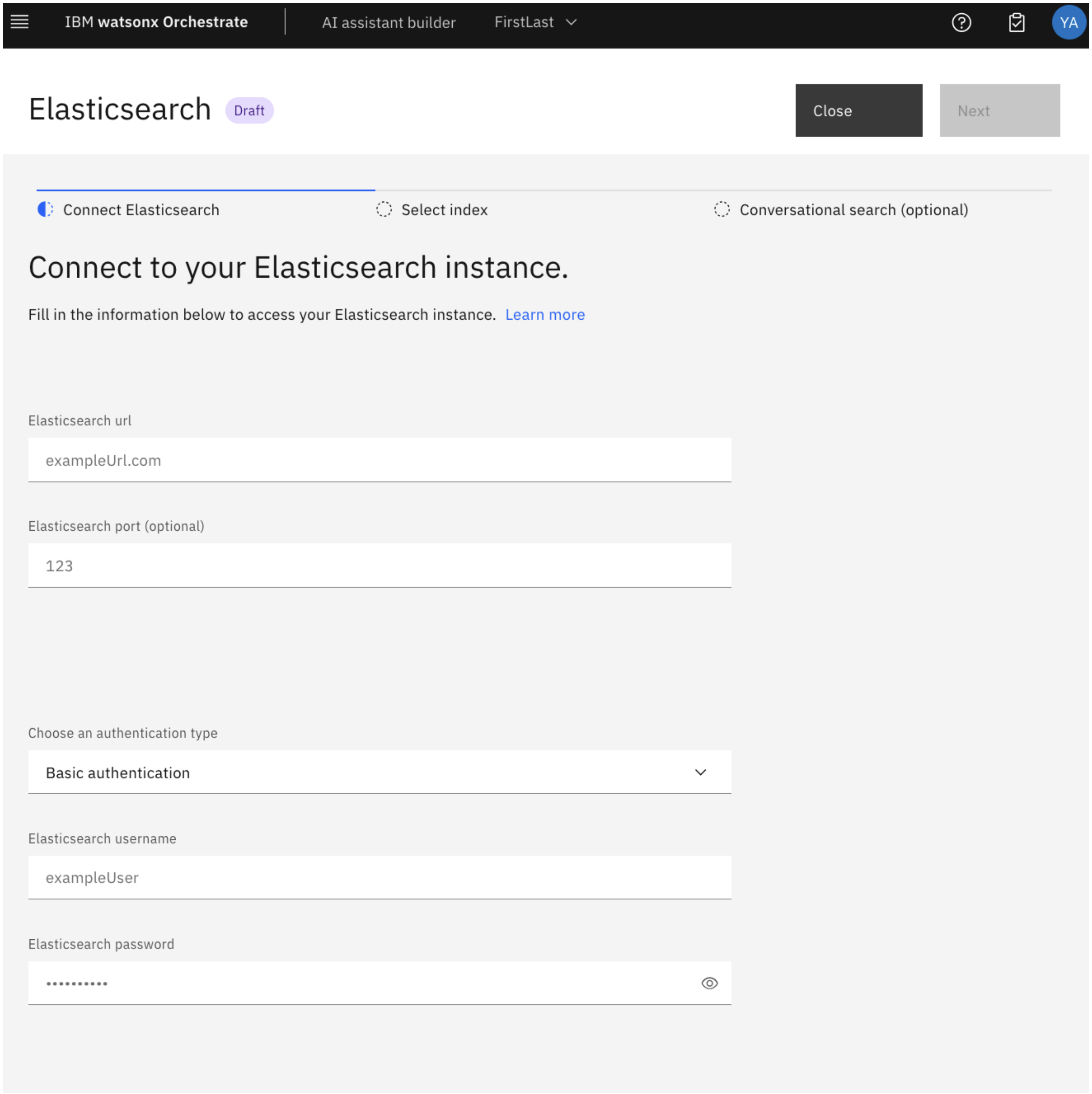
Custom service

Connect to an existing service that will retrieve documents from your content management service.Learn more

Note: you can also connect to watsonx.data Milvus instead of Elasticsearch – follow this guide.

5. You should now see the following screen prompting you to enter **Elastcisearch** credentials:

© 2024 IBM Corporation



Connect to your Elasticsearch instance.

Fill in the information below to access your Elasticsearch instance. Learn more

Elasticsearch url: exampleUrl.com

Elasticsearch port (optional): 123

Choose an authentication type: Basic authentication

Elasticsearch username: exampleUser

Elasticsearch password: ••••••••••

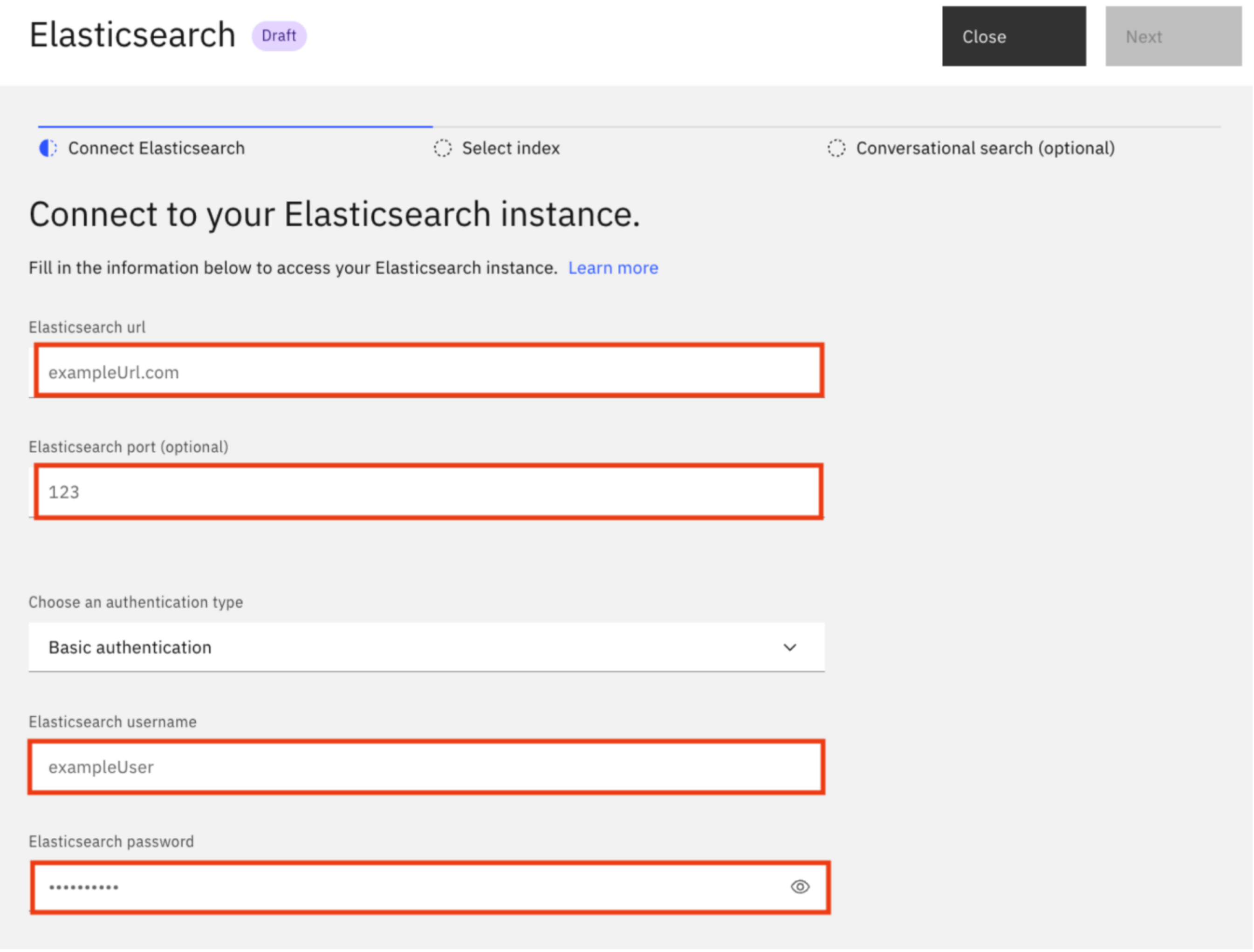
6. Connect to the Elasticsearch instance in watsonx Orchestrate using the following values provided by your proctor:

   Elasticsearch url (make sure to remove the trailing slash, if any)

   Elasticsearch port

© 2024 IBM Corporation

Leave authentication type as **Basic authentication** and fill in the **Elasticsearch username** and **Elasticsearch password**



Connect to your Elasticsearch instance.

Fill in the information below to access your Elasticsearch instance. Learn more

Elasticsearch url: exampleUrl.com

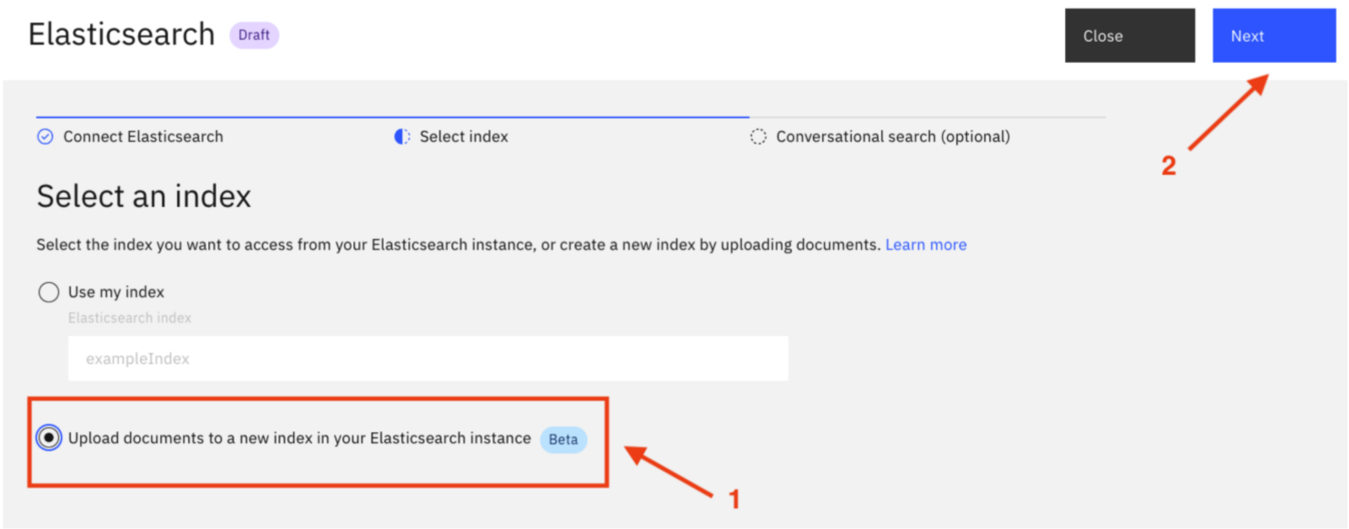
Elasticsearch port (optional): 123

Choose an authentication type: Basic authentication

Elasticsearch username: exampleUser

Elasticsearch password: ••••••••••

7. Click **Next**
8. On the next screen select **Upload documents to a new index in your Elasticsearch instance** to automatically create a new index in Elasticsearch and click **Next** to continue



Select an index

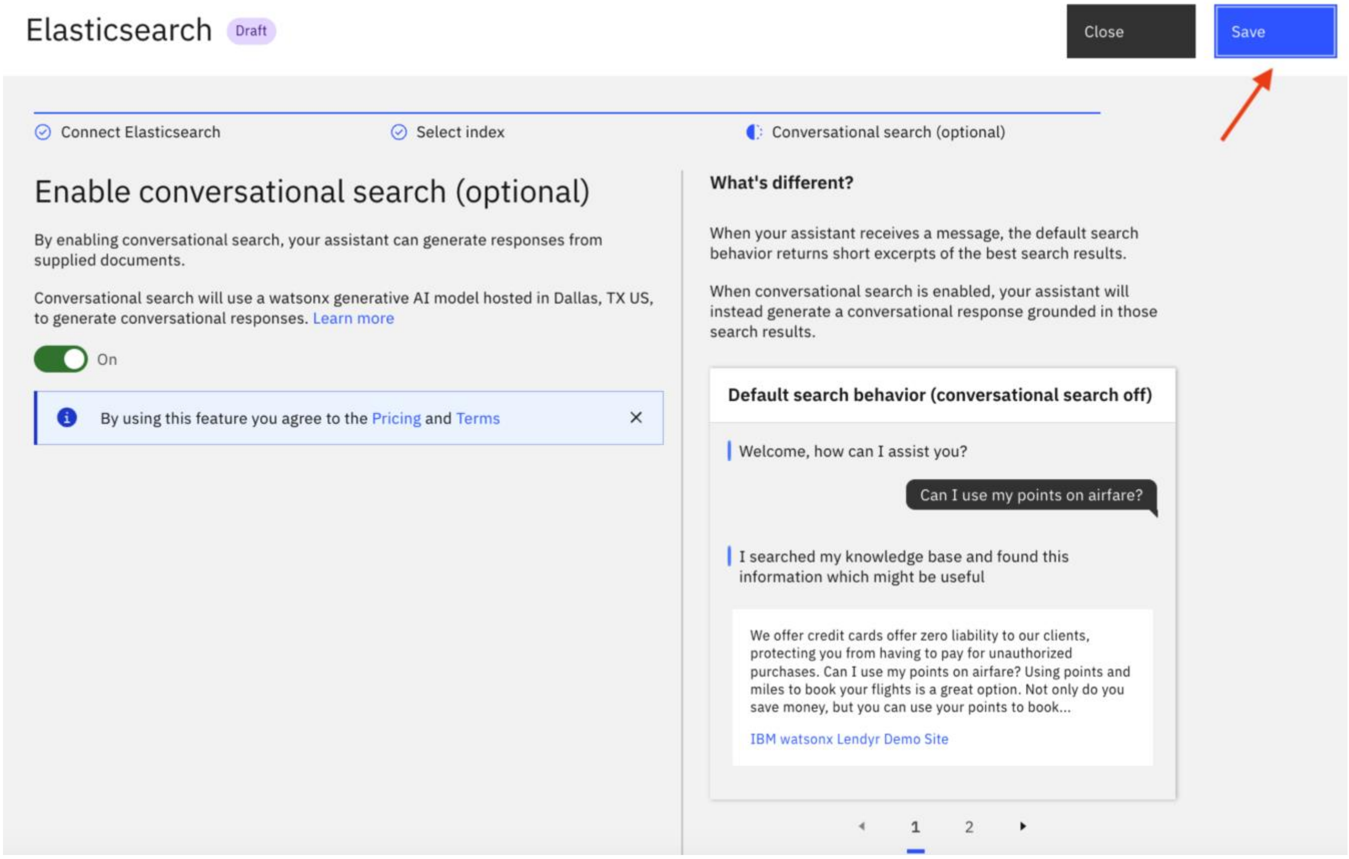
Select the index you want to access from your Elasticsearch instance, or create a new index by uploading documents. Learn more

○ Use my index

Elasticsearch index: exampleIndex

● Upload documents to a new index in your Elasticsearch instance Beta

9. On the next screen make sure conversational search is **On** and click **Save**



Enable conversational search (optional)

By enabling conversational search, your assistant can generate responses from supplied documents.

Conversational search will use a watsonx generative AI model hosted in Dallas, TX US, to generate conversational responses. Learn more

On

By using this feature you agree to the Pricing and Terms

What's different?

When your assistant receives a message, the default search behavior returns short excerpts of the best search results.

When conversational search is enabled, your assistant will instead generate a conversational response grounded in those search results.

Default search behavior (conversational search off)

Welcome, how can I assist you?

Can I use my points on airfare?

I searched my knowledge base and found this information which might be useful

We offer credit cards offer zero liability to our clients, protecting you from having to pay for unauthorized purchases. Can I use my points on airfare? Using points and miles to book your flights is a great option. Not only do you save money, but you can use your points to book...

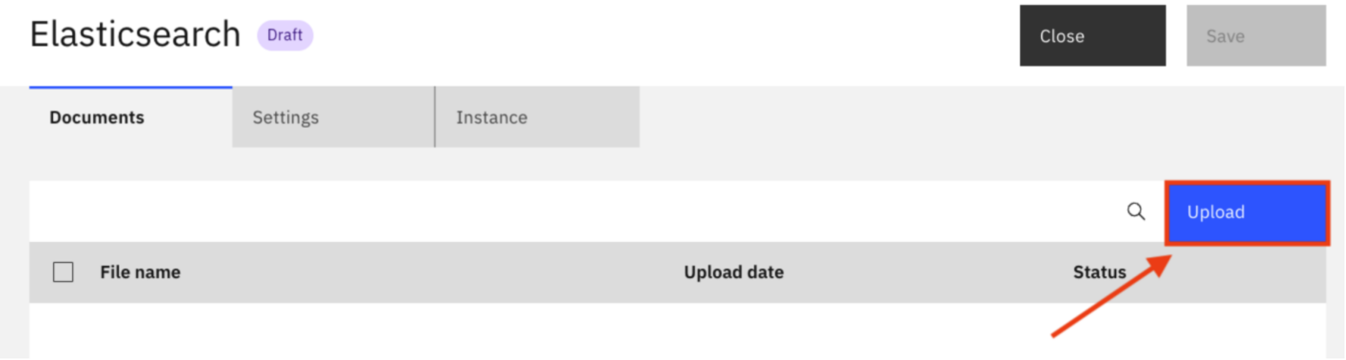
IBM watsonx Lendyr Demo Site

Once you click **Save,** you will be taken to the **Documents** tab where you can upload your documents to the knowledge base (see next section).

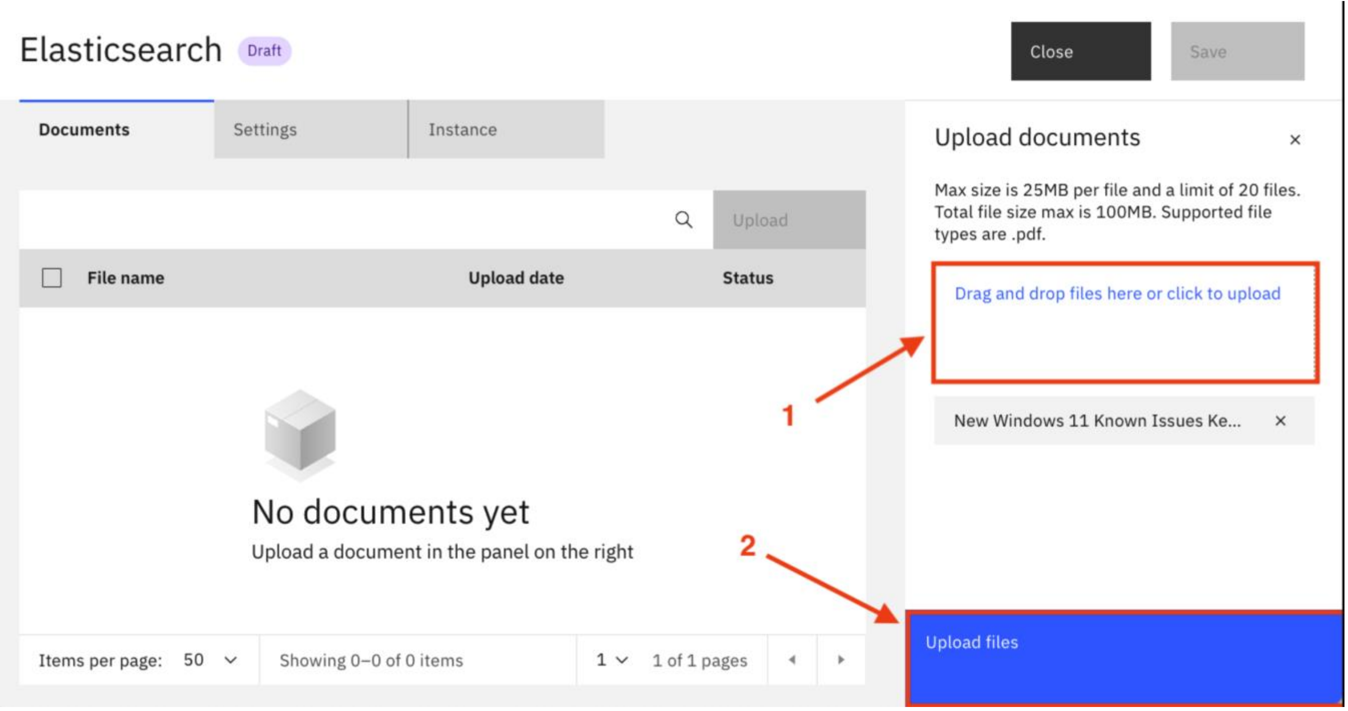
© 2024 IBM Corporation

Upload documents to the knowledge base

1. Now we need to upload our documents into the knowledge base. In the **Documents** tab click **Upload**



2. Find the file **New Windows 11 Known Issues Keep Coming.pdf** (provided by your proctor) and drag and drop the .pdf file into the drag and drop area. Click on **Upload files**



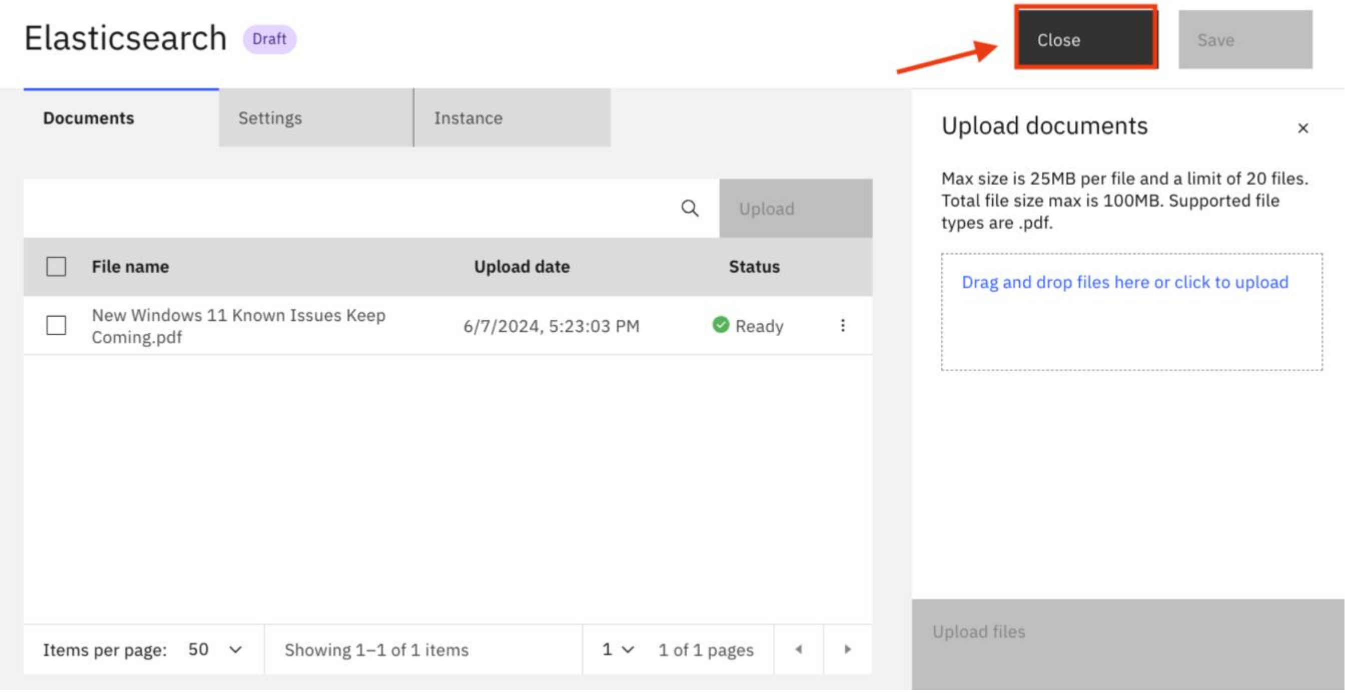
No documents yet

Upload a document in the panel on the right

Max size is 25MB per file and a limit of 20 files. Total file size max is 100MB. Supported file types are .pdf.

3. Once the document is ready, the screen will look the following way and you can click on **Close**

© 2024 IBM Corporation



☐	File name	Upload date	Status
☐	New Windows 11 Known Issues Keep Coming.pdf	6/7/2024, 5:23:03 PM	Ready

Items per page: 50

Showing 1–1 of 1 items

1 of 1 pages

Max size is 25MB per file and a limit of 20 files. Total file size max is 100MB. Supported file types are .pdf.

Drag and drop files here or click to upload

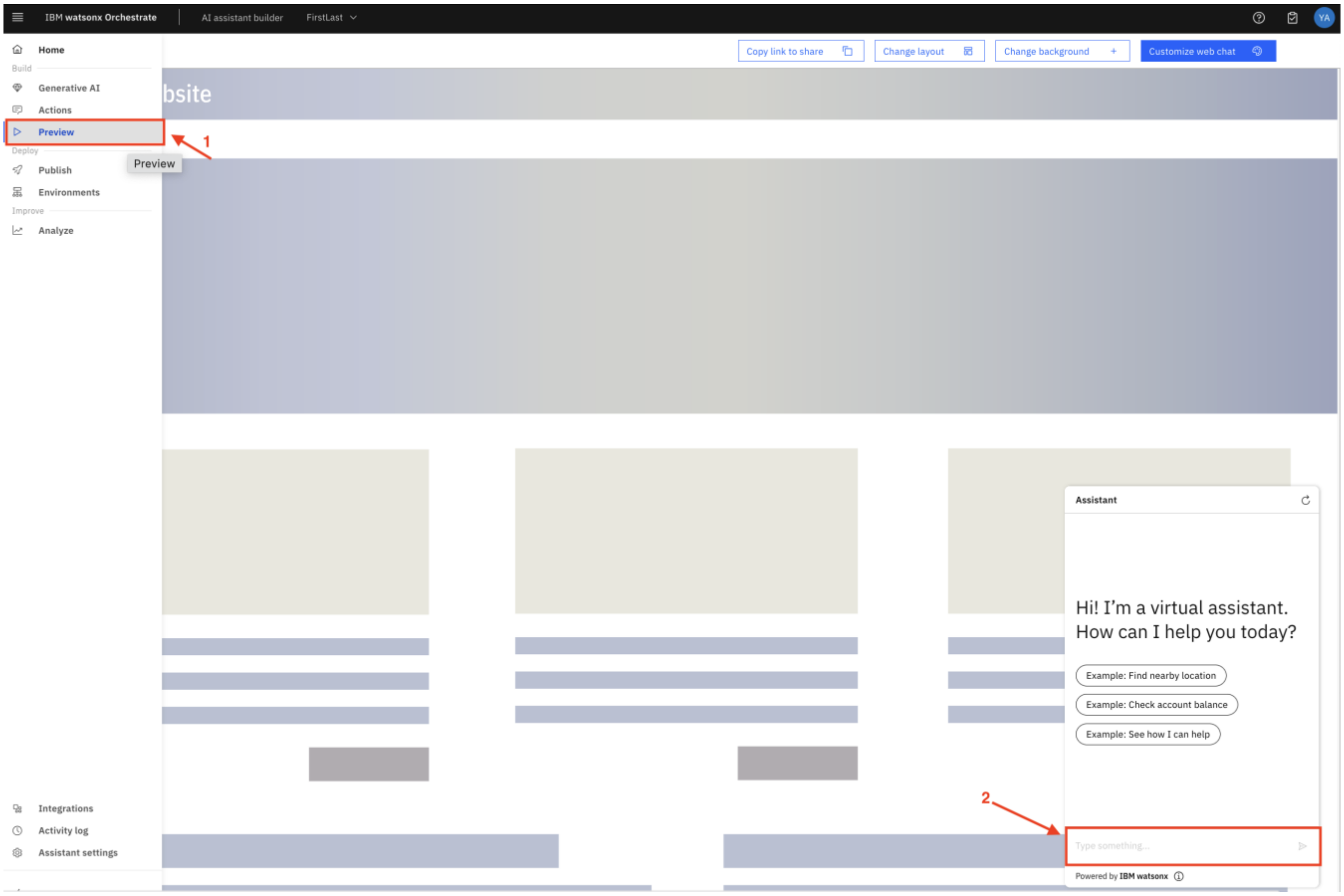
And this is it. You have now finished the process of ingesting documents into the **Elasticsearch** knowledge base!

Test the virtual assistant

Now that conversational search is set up and the document has been uploaded to the knowledge base, we can test the virtual assistant by asking some questions 1) based on the Q&As in the document 2) general questions that may be answered by a built in LLM.

1. Open **Preview** and go into the input window in the virtual assistant:

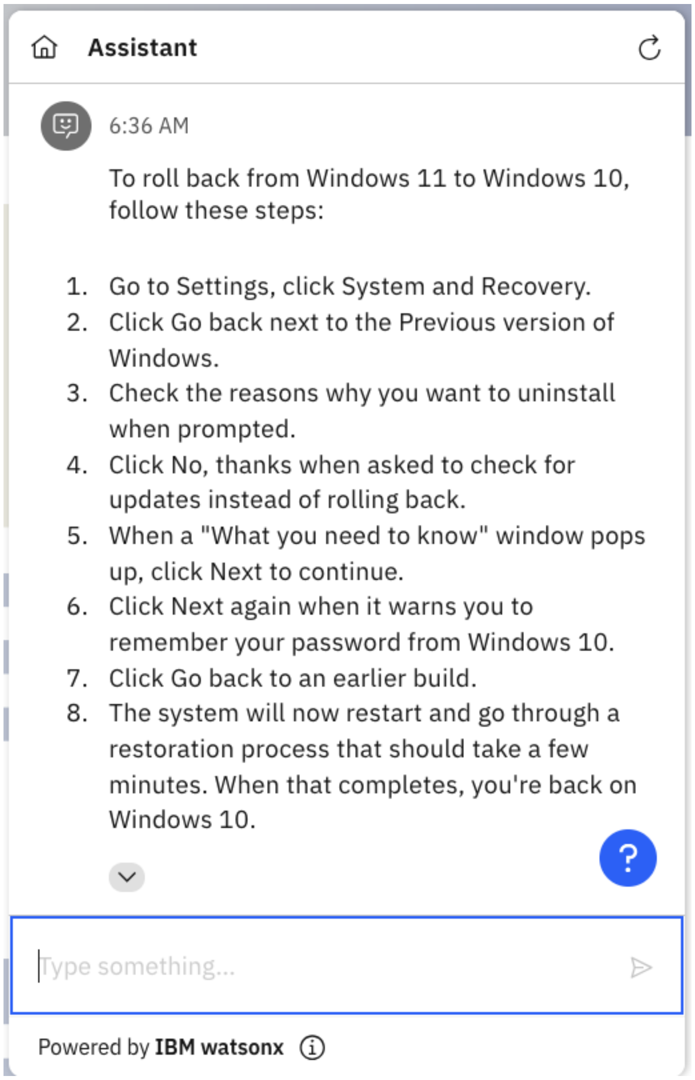
© 2024 IBM Corporation



Hi! I'm a virtual assistant. How can I help you today?

2. Type a question that is answered in the document, e.g. "how do I roll back from windows 11 to windows 10?" and check that the answer provided was retrieved from the document:

© 2024 IBM Corporation



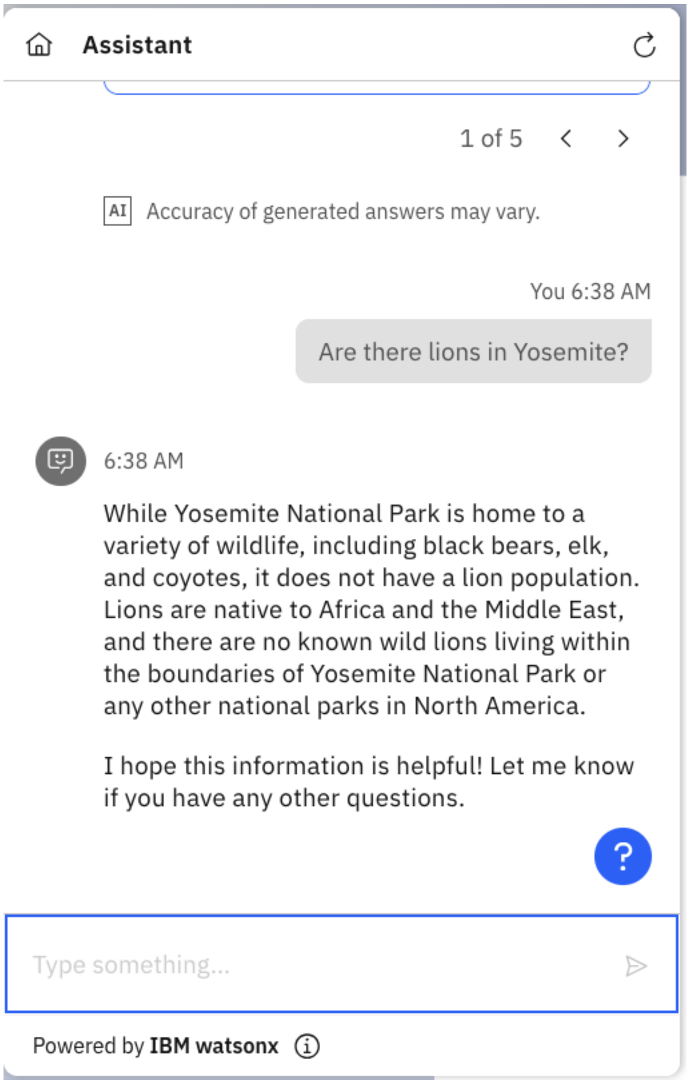
6:36 AM

To roll back from Windows 11 to Windows 10, follow these steps:

1. Go to Settings, click System and Recovery.
2. Click Go back next to the Previous version of Windows.
3. Check the reasons why you want to uninstall when prompted.
4. Click No, thanks when asked to check for updates instead of rolling back.
5. When a "What you need to know" window pops up, click Next to continue.
6. Click Next again when it warns you to remember your password from Windows 10.
7. Click Go back to an earlier build.
8. The system will now restart and go through a restoration process that should take a few minutes. When that completes, you're back on Windows 10.

3. Type a question that is NOT answered in the document, e.g. "Are there lions in Yosemite?" and check if the LLM provides an answer

© 2024 IBM Corporation



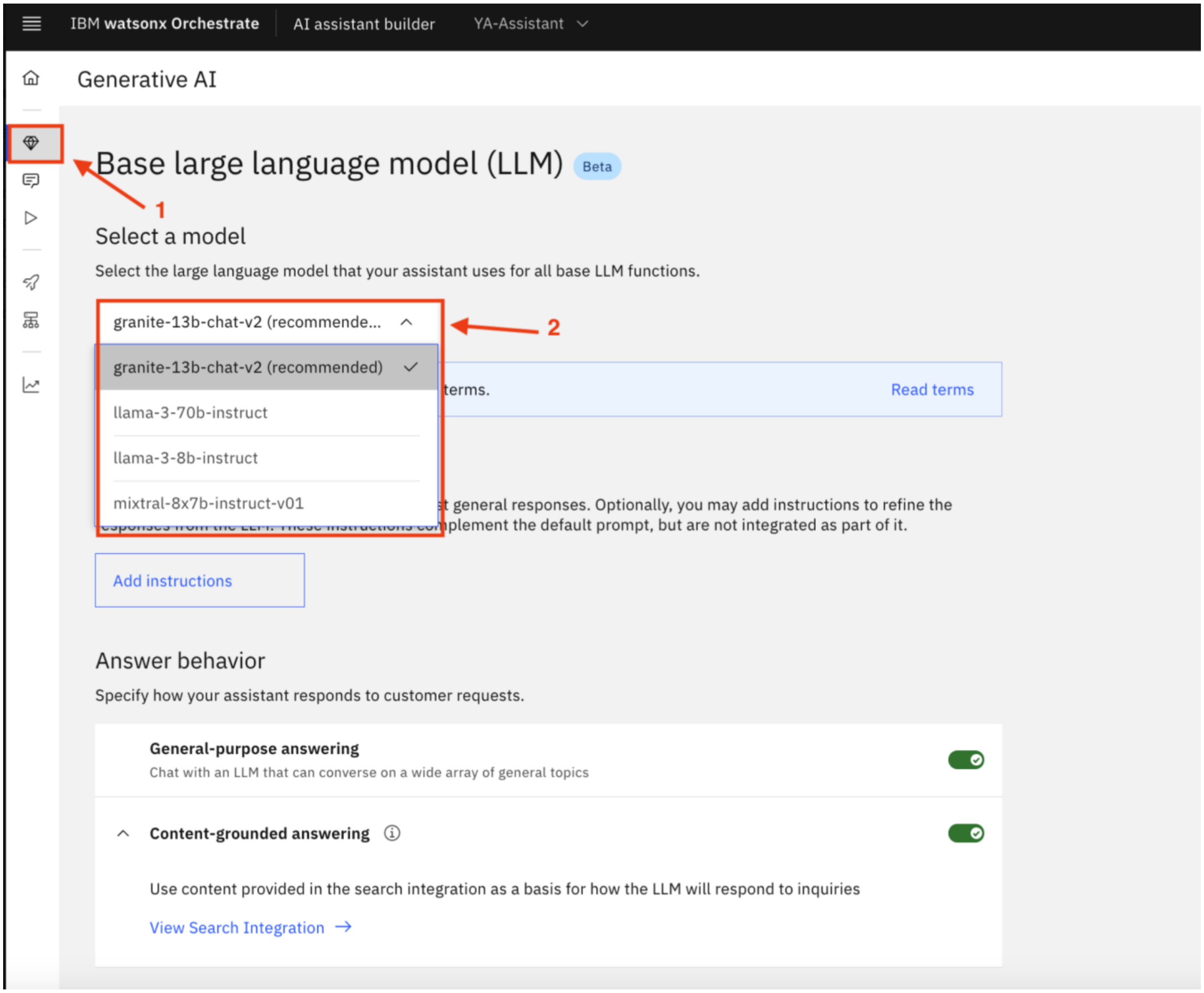
Are there lions in Yosemite?

While Yosemite National Park is home to a variety of wildlife, including black bears, elk, and coyotes, it does not have a lion population. Lions are native to Africa and the Middle East, and there are no known wild lions living within the boundaries of Yosemite National Park or any other national parks in North America.

I hope this information is helpful! Let me know if you have any other questions.

4. **(Optional)** You can test out different LLMs available under the **Generative AI** tab and observe how different the answers are to the same question

© 2024 IBM Corporation



Select a model

Select the large language model that your assistant uses for all base LLM functions.

Answer behavior

Specify how your assistant responds to customer requests.

General-purpose answering

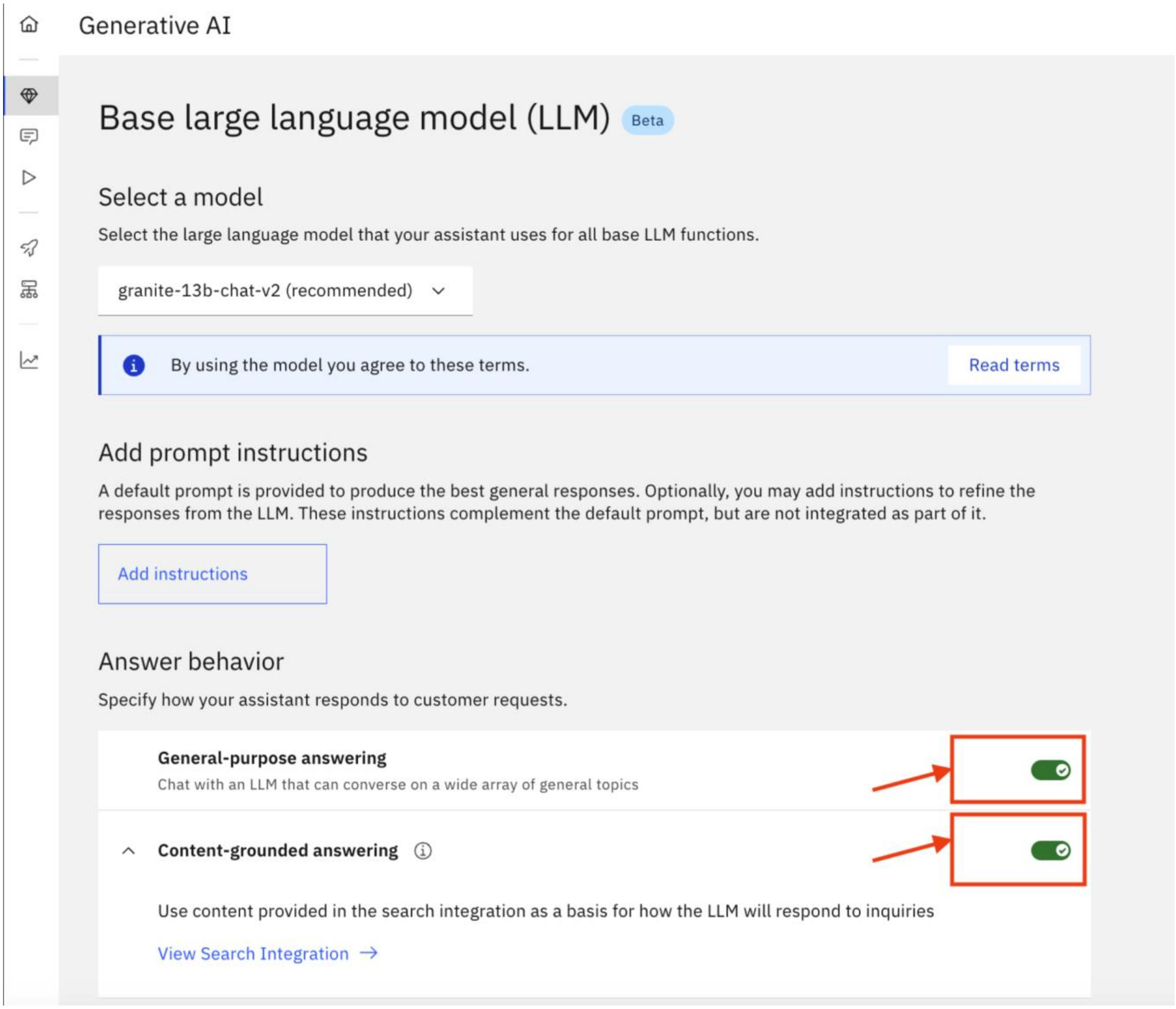
Chat with an LLM that can converse on a wide array of general topics

Content-grounded answering

Use content provided in the search integration as a basis for how the LLM will respond to inquiries

5. **(Optional)** You can also experiment with the answer behavior by toggling **General-purpose answering** and **Content-grounded answering** options on and off. See how the answers change with different combinations of these settings.

© 2024 IBM Corporation



Generative AI

Base large language model (LLM) Beta

Select a model

Select the large language model that your assistant uses for all base LLM functions.

granite-13b-chat-v2 (recommended)

By using the model you agree to these terms. Read terms

Add prompt instructions

A default prompt is provided to produce the best general responses. Optionally, you may add instructions to refine the responses from the LLM. These instructions complement the default prompt, but are not integrated as part of it.

Add instructions

Answer behavior

Specify how your assistant responds to customer requests.

General-purpose answering

Chat with an LLM that can converse on a wide array of general topics

Content-grounded answering

Use content provided in the search integration as a basis for how the LLM will respond to inquiries

View Search Integration →

Note: Given the timeframe available for this hands-on lab, we experimented with only one .pdf file. However, you could upload multiple files and ask questions about information contained in any of them. The answers provided by conversational search will be reference the original source.

In the steps above you configured conversational search in your assistant with a knowledge base that stores FAQs (in Elasticsearch). When the user asks a question, the knowledge base is queried to retrieve any relevant FAQs which are then passed to a built-in watsonx.ai LLM (IBM Granite) to generate an answer for the customer.

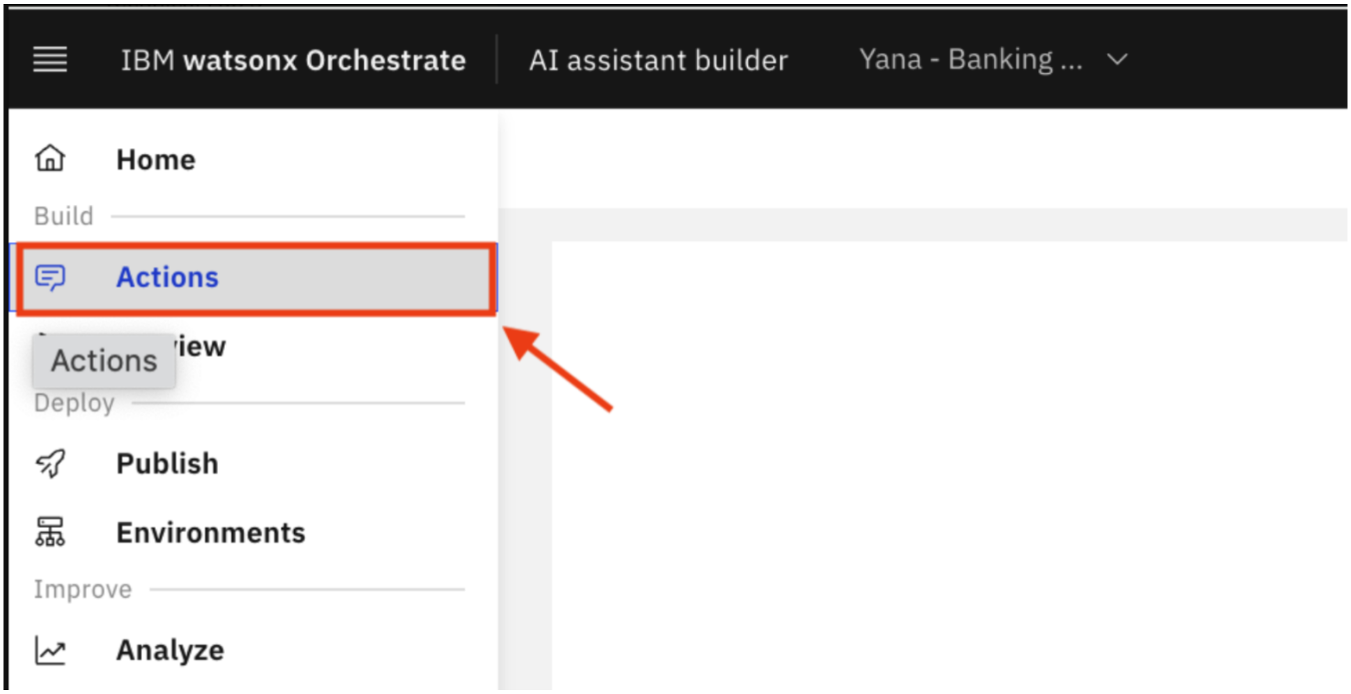
© 2024 IBM Corporation

Implement an AI guided action

Create an AI-guided action to answer questions

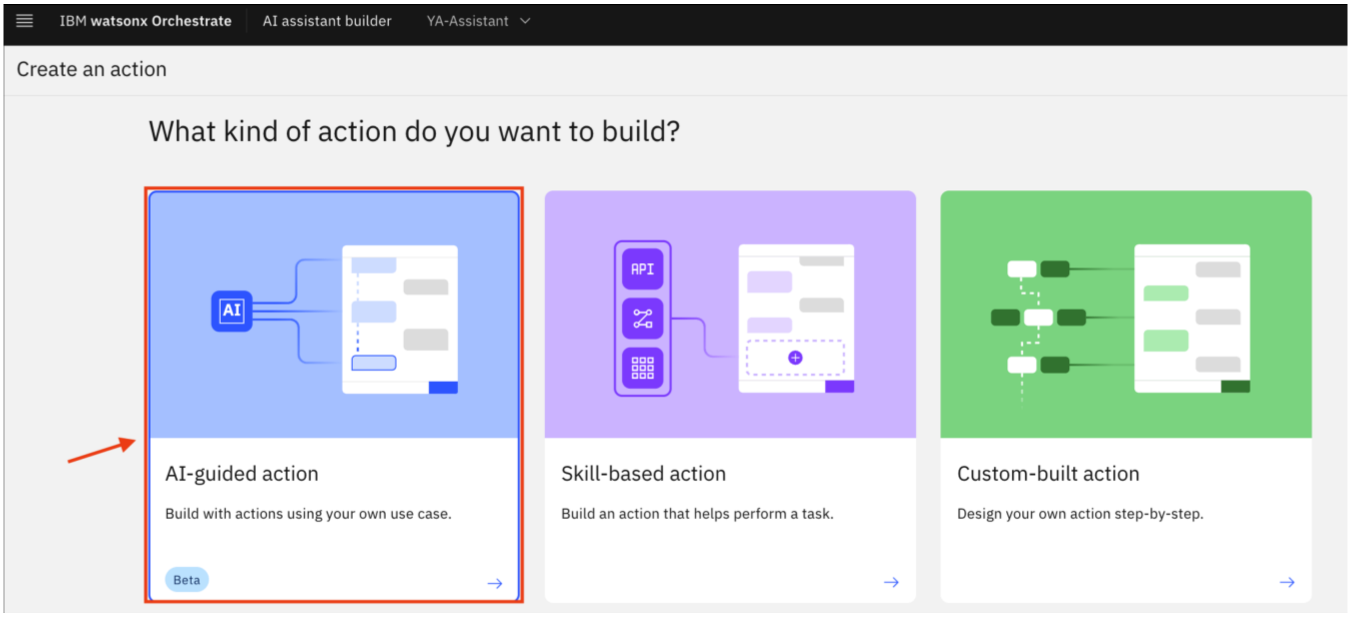
In this step we will create an **AI-guided action** to answer questions from a user related to a specific subject, given provided knowledge context. This is different from conversational search in that the answers are grounded in very specific knowledge. At the same time, the answers to AI-guided actions are limited to the specific knowledge provided vs. the **general-purpose answering** where LLMs are allowed to answer the user's question based on all the vast data they have been trained on.

6. Go into your assistant builder instance (make sure your name is selected in the drop-down) and select the **Actions** tab



7. Create a new action and select **AI-guided action**

© 2024 IBM Corporation



Create an action

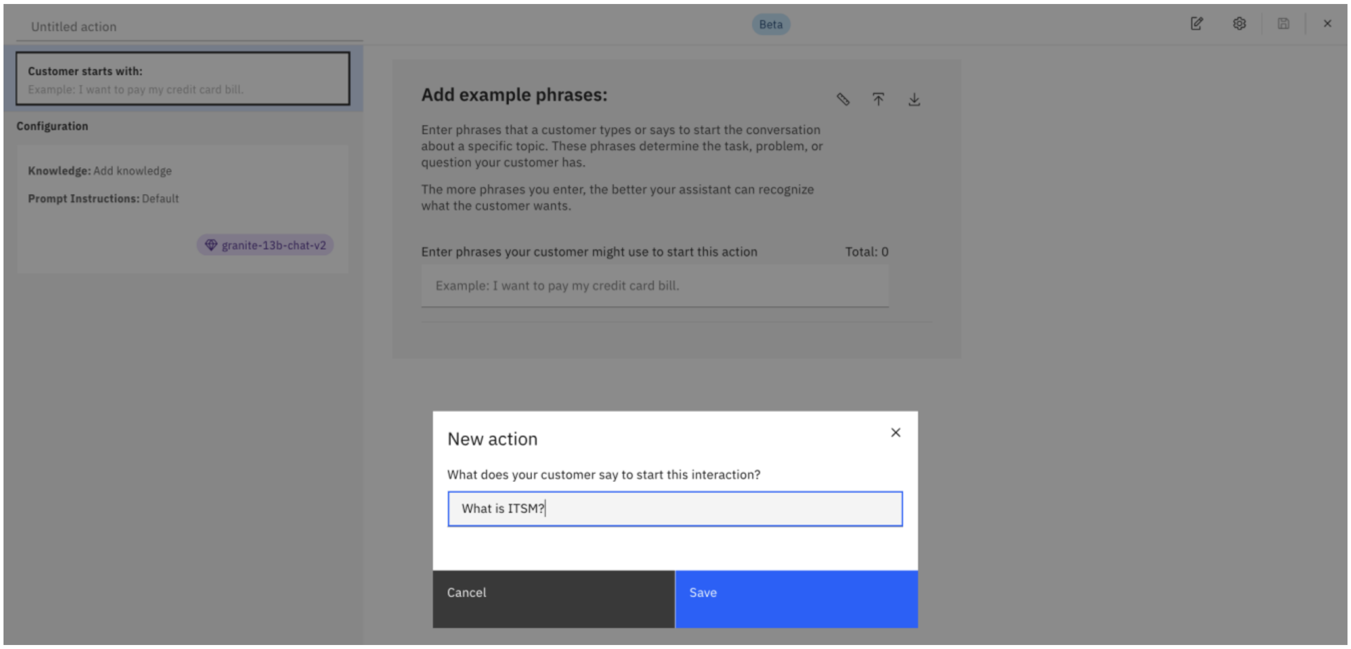
What kind of action do you want to build?

AI-guided action
Build with actions using your own use case.

Skill-based action
Build an action that helps perform a task.

Custom-built action
Design your own action step-by-step.

8. Pick a topic that the user may want to ask about, e.g. IT Service Management (ITSM) and provide a phrase that a user might type that is related to this subject, e.g. "What is ITSM"



Customer starts with:
Example: I want to pay my credit card bill.

Knowledge: Add knowledge
Prompt Instructions: Default

Add example phrases:

Enter phrases that a customer types or says to start the conversation about a specific topic. These phrases determine the task, problem, or question your customer has.

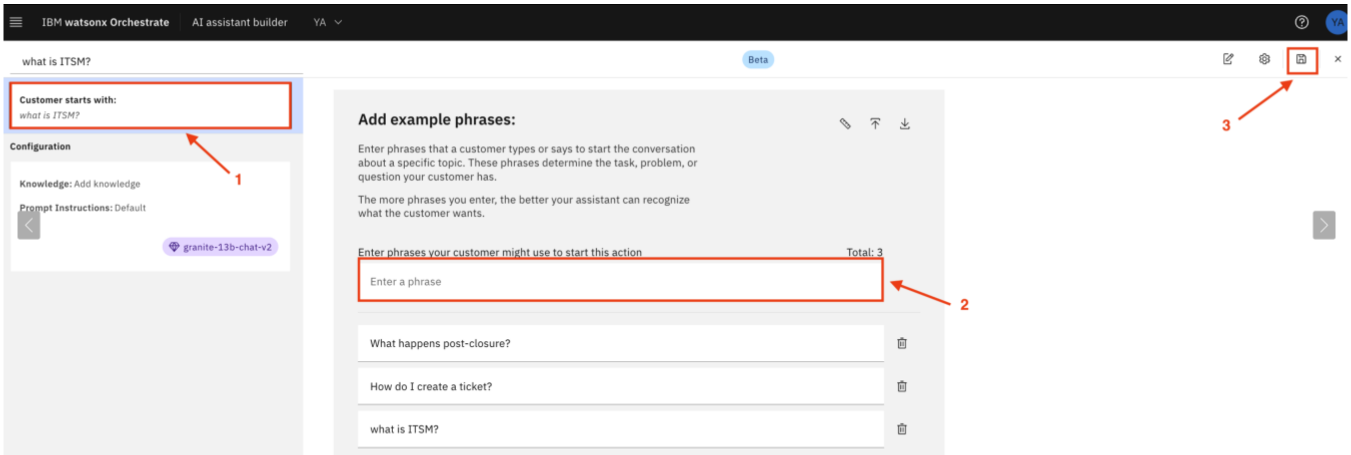
The more phrases you enter, the better your assistant can recognize what the customer wants.

Enter phrases your customer might use to start this action — Total: 0
Example: I want to pay my credit card bill.

New action
What does your customer say to start this interaction?
What is ITSM?

9. Go into **Customer starts with:** in the action configuration menu and provide additional phrases/words that the user might type to ask about this subject, each time hitting Enter, and then **Save** everything

© 2024 IBM Corporation



Add example phrases:

Enter phrases that a customer types or says to start the conversation about a specific topic. These phrases determine the task, problem, or question your customer has.

The more phrases you enter, the better your assistant can recognize what the customer wants.

Enter phrases your customer might use to start this action Total: 3

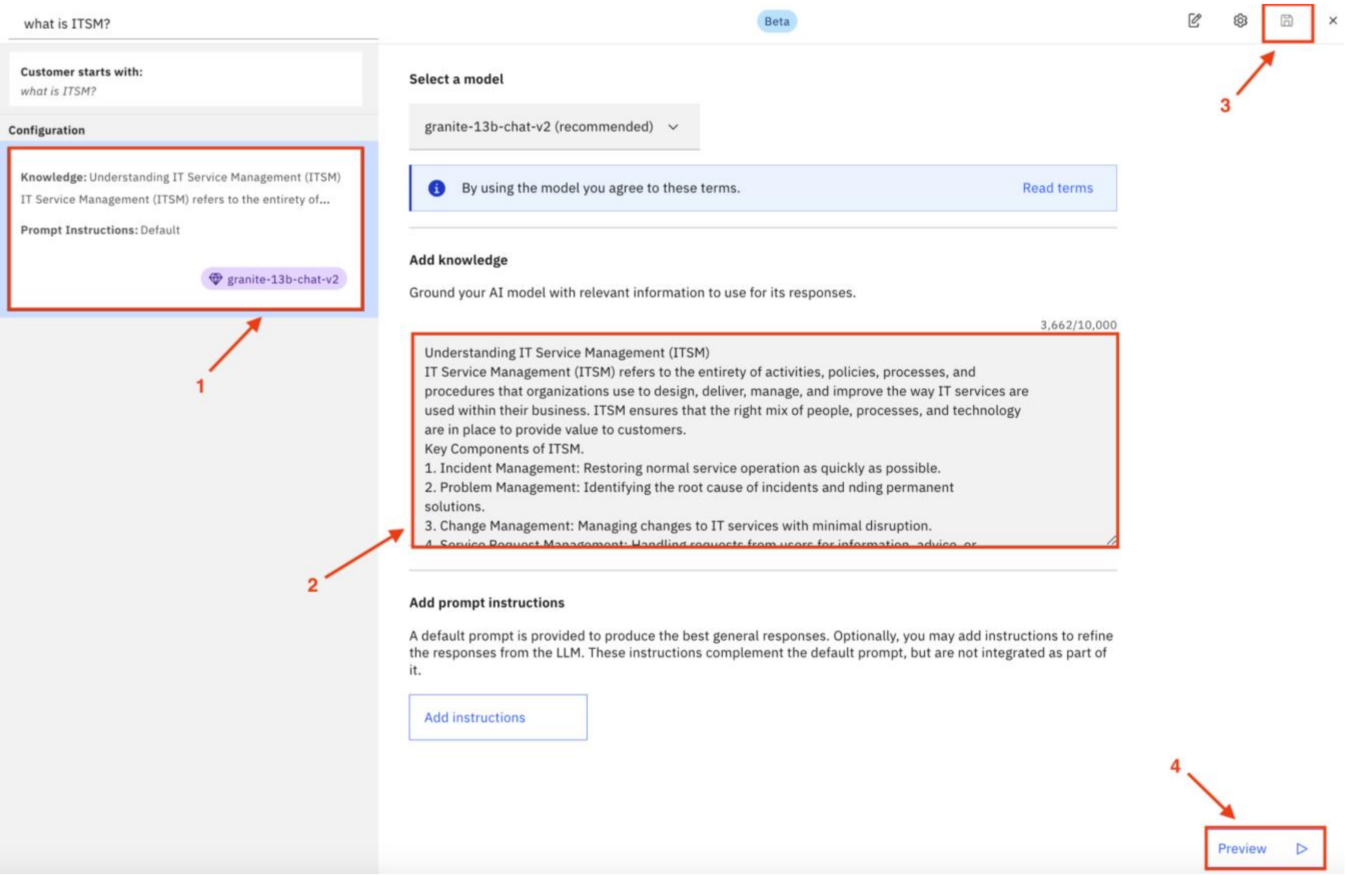
Enter a phrase

What happens post-closure?

How do I create a ticket?

what is ITSM?

10. Identify related content on this subject and provide it as **knowledge** inside the action (you can use the text below the following screenshot as a sample). Make sure to save again and then click on the preview button in the bottom right of the screen to test your action


what is ITSM?
Beta
Customer starts with:
what is ITSM?
Configuration
Knowledge: Understanding IT Service Management (ITSM) IT Service Management (ITSM) refers to the entirety of...
Prompt Instructions: Default
granite-13b-chat-v2
1
2
3
4
Read terms
Add instructions
Preview

Select a model

granite-13b-chat-v2 (recommended)

By using the model you agree to these terms.

Add knowledge

Ground your AI model with relevant information to use for its responses.

3,662/10,000

Understanding IT Service Management (ITSM)
IT Service Management (ITSM) refers to the entirety of activities, policies, processes, and procedures that organizations use to design, deliver, manage, and improve the way IT services are used within their business. ITSM ensures that the right mix of people, processes, and technology are in place to provide value to customers.
Key Components of ITSM.
1. Incident Management: Restoring normal service operation as quickly as possible.
2. Problem Management: Identifying the root cause of incidents and nding permanent solutions.
3. Change Management: Managing changes to IT services with minimal disruption.
4. Service Request Management: Handling requests from users for information, advice, or

Add prompt instructions

A default prompt is provided to produce the best general responses. Optionally, you may add instructions to refine the responses from the LLM. These instructions complement the default prompt, but are not integrated as part of it.

Text that can be used as sample knowledge:

Understanding IT Service Management (ITSM)

IT Service Management (ITSM) refers to the entirety of activities, policies, processes, and procedures that organizations use to design, deliver, manage, and improve the way IT services are used within their business. ITSM ensures that the right mix of people, processes, and technology are in place to provide value to customers.

Key Components of ITSM

1. **Incident Management**: Restoring normal service operation as quickly as possible.
2. **Problem Management**: Identifying the root cause of incidents and finding permanent solutions.
3. **Change Management**: Managing changes to IT services with minimal disruption.
4. **Service Request Management**: Handling requests from users for information, advice, or access to IT services.
5. **Asset and Configuration Management**: Keeping track of IT assets and configurations.
6. **Knowledge Management**: Creating, sharing, using, and managing knowledge and information.

ITSM Workflows for Creating a Ticket

Creating a ticket in ITSM involves several steps. Here's a generalized workflow:

1. **Ticket Creation**
    - **User Submission**: A user submits a ticket via an online portal, email, or phone.
    - **Automatic Creation**: The system may create tickets automatically through monitoring tools.
2. **Categorization**
    - **Ticket Classification**: Categorize the ticket based on the type of issue (e.g., incident, service request).
    - **Priority Assignment**: Assign priority based on the impact and urgency of the issue.
3. **Assignment**
    - **Routing**: Route the ticket to the appropriate support group or technician.
    - **Automatic Assignment**: The system may automatically assign tickets based on predefined rules.

© 2024 IBM Corporation

4. **Investigation and Diagnosis**
    - **Initial Assessment**: The assigned technician reviews the ticket details.
    - **Diagnosis**: Investigate the issue and diagnose the root cause.
    - **Update Ticket**: Update the ticket with diagnostic information.
5. **Resolution and Recovery**
    - **Fix Implementation**: Implement the solution or workaround.
    - **User Communication**: Communicate the solution to the user and confirm resolution.
    - **Update Ticket**: Document the steps taken to resolve the issue.
6. **Closure**
    - **User Confirmation**: Confirm with the user that the issue is resolved satisfactorily.
    - **Ticket Closure**: Close the ticket and update the system.
7. **Post-Closure Review**
    - **Review and Analysis**: Conduct a review to analyze the issue and identify improvements.
    - **Knowledge Base Update**: Update the knowledge base with new insights or solutions.

Example of an ITSM Ticket Workflow

1. **Step 1**: Ticket Creation
    - **User**: John Doe reports an issue with email access through the IT service portal.
2. **Step 2**: Categorization
    - **System**: Categorizes the ticket as an "Email Issue" with a high priority.
3. **Step 3**: Assignment
    - **System**: Automatically assigns the ticket to the "Email Support" team based on predefined rules.
4. **Step 4**: Investigation and Diagnosis
    - **Technician**: Jane Smith reviews the ticket and identifies a server issue as the root cause.

© 2024 IBM Corporation

    - **Technician**: Documents her findings and the steps she will take to resolve the issue.

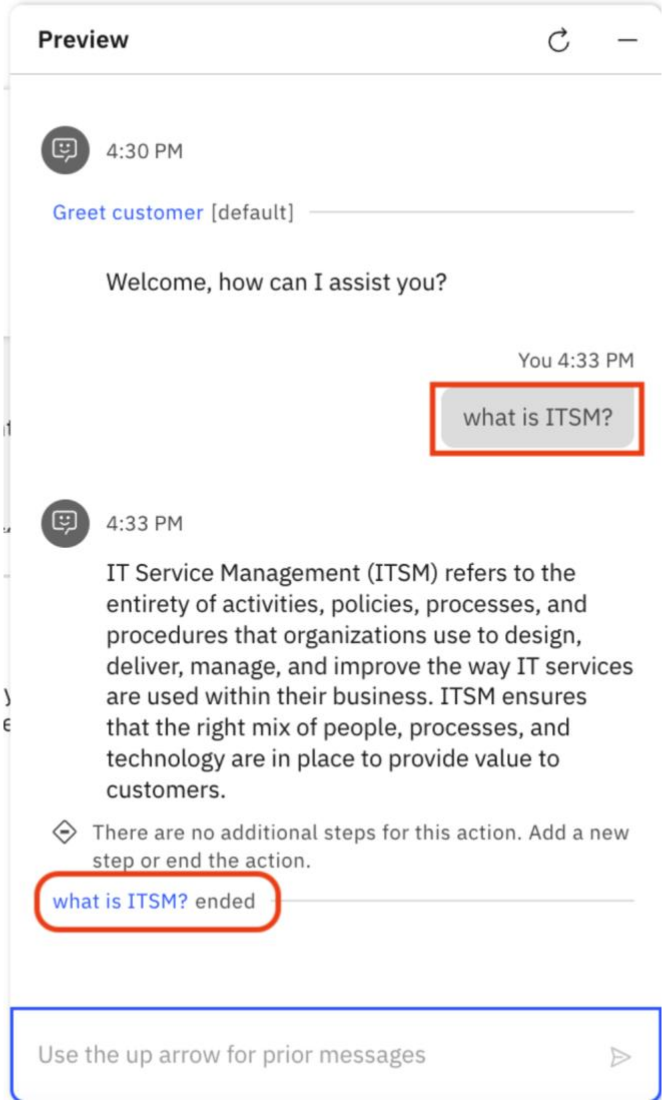
5. **Step 5**: Resolution and Recovery
    - **Technician**: Implements the fix by restarting the email server.
    - **Technician**: Notifies John Doe that the issue has been resolved and asks him to verify.
6. **Step 6**: Closure
    - **User**: John Doe confirms that the email access issue is resolved.
    - **Technician**: Closes the ticket and updates the resolution details in the system.
7. **Step 7**: Post-Closure Review
    - **Team Lead**: Conducts a review of the incident to identify any improvements or patterns.
8. **Knowledge Base**: Updates the knowledge base with the resolution steps for future reference.

This workflow ensures efficient handling of issues, leading to better service quality and user satisfaction.

Test your AI guided action

1. Test the action by typing some of the sample phrases/words you provided earlier and make sure your new action was invoked and the results are based on the knowledge you provided:

© 2024 IBM Corporation



4:30 PM

Greet customer [default]

Welcome, how can I assist you?

You 4:33 PM

what is ITSM?

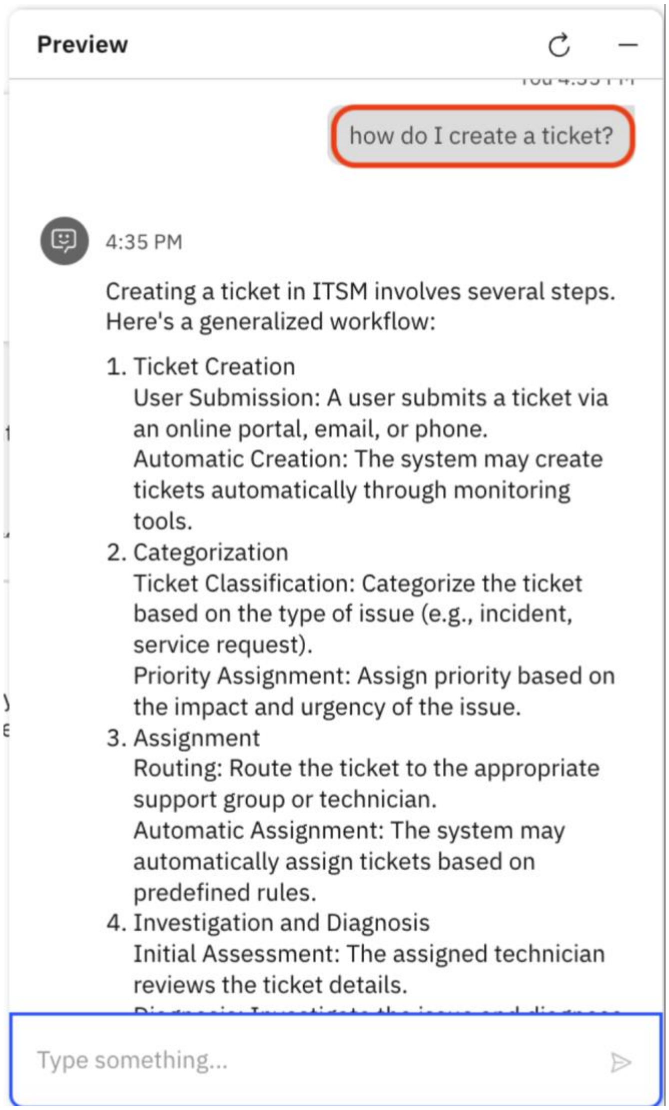
4:33 PM

IT Service Management (ITSM) refers to the entirety of activities, policies, processes, and procedures that organizations use to design, deliver, manage, and improve the way IT services are used within their business. ITSM ensures that the right mix of people, processes, and technology are in place to provide value to customers.

There are no additional steps for this action. Add a new step or end the action.

what is ITSM? ended

© 2024 IBM Corporation



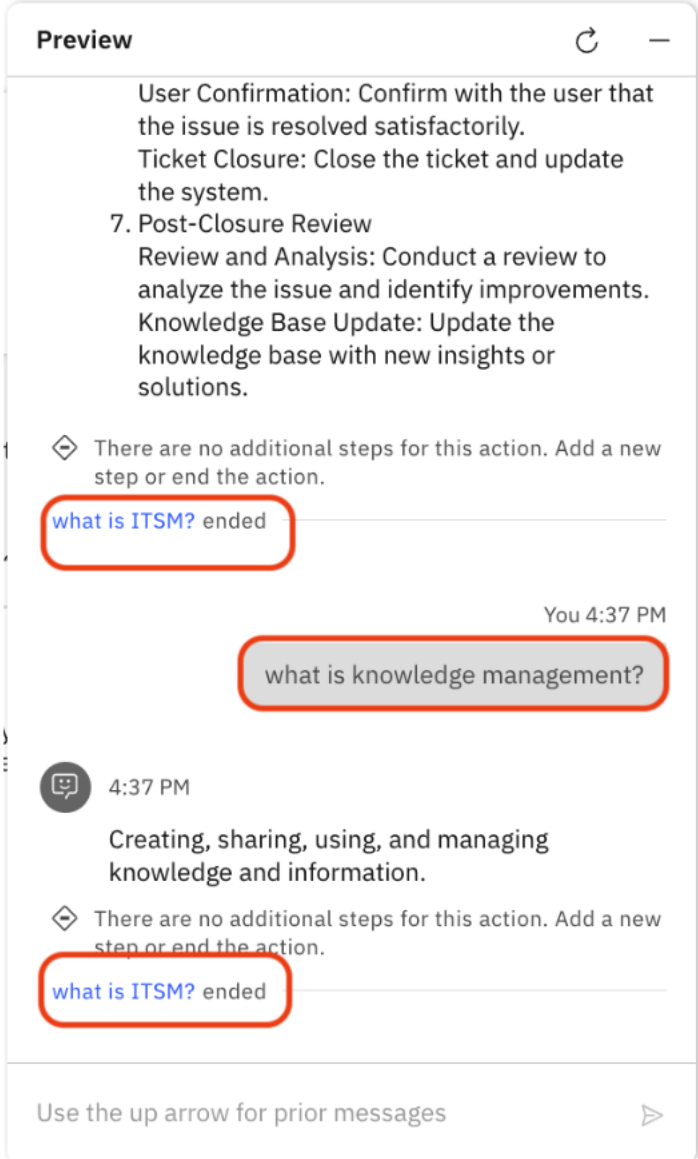
how do I create a ticket?

4:35 PM

Creating a ticket in ITSM involves several steps. Here's a generalized workflow:

1. Ticket Creation
   User Submission: A user submits a ticket via an online portal, email, or phone.
   Automatic Creation: The system may create tickets automatically through monitoring tools.
2. Categorization
   Ticket Classification: Categorize the ticket based on the type of issue (e.g., incident, service request).
   Priority Assignment: Assign priority based on the impact and urgency of the issue.
3. Assignment
   Routing: Route the ticket to the appropriate support group or technician.
   Automatic Assignment: The system may automatically assign tickets based on predefined rules.
4. Investigation and Diagnosis
   Initial Assessment: The assigned technician reviews the ticket details.

© 2024 IBM Corporation



User Confirmation: Confirm with the user that the issue is resolved satisfactorily.
Ticket Closure: Close the ticket and update the system.

7. Post-Closure Review
Review and Analysis: Conduct a review to analyze the issue and identify improvements.
Knowledge Base Update: Update the knowledge base with new insights or solutions.

There are no additional steps for this action. Add a new step or end the action.

what is ITSM? ended

You 4:37 PM

what is knowledge management?

4:37 PM

Creating, sharing, using, and managing knowledge and information.

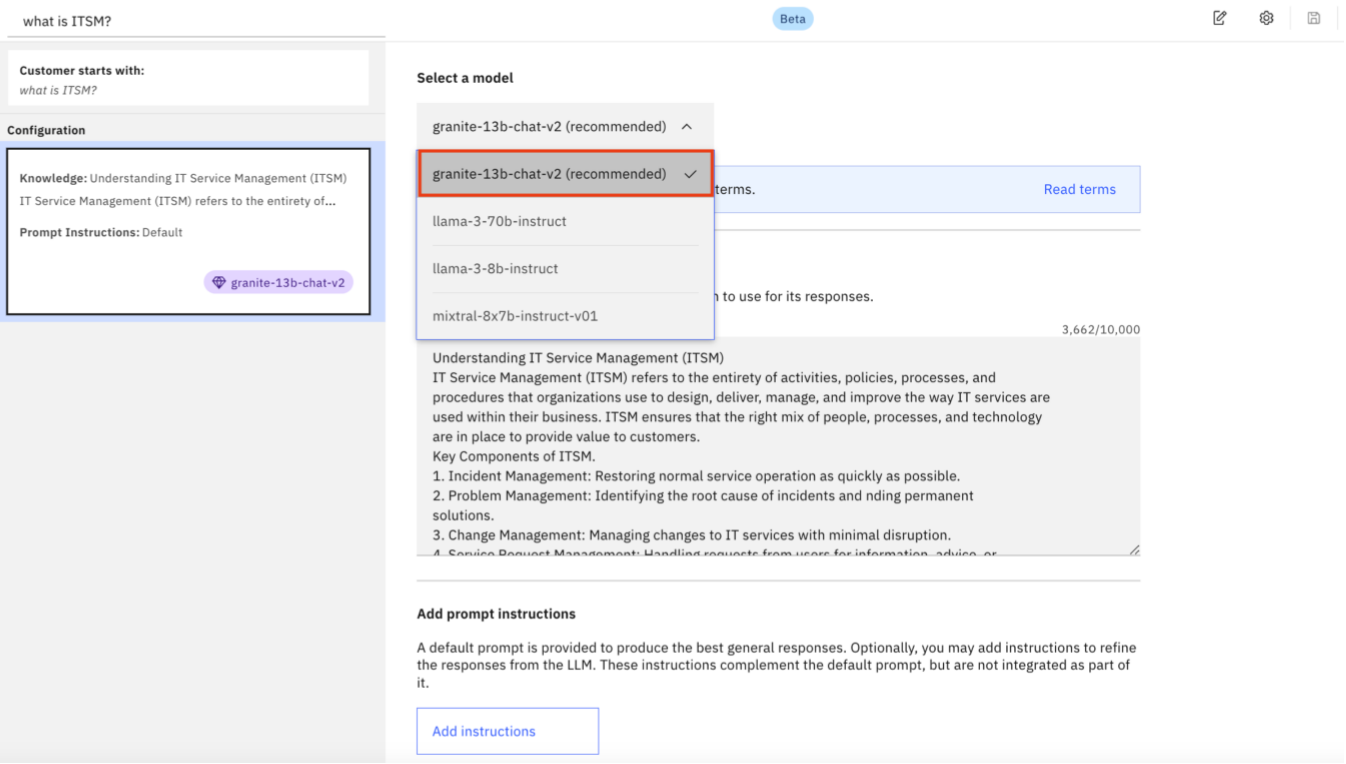
There are no additional steps for this action. Add a new step or end the action.

what is ITSM? ended

Use the up arrow for prior messages

2. Feel free to experiment with additional settings, e.g. selecting a different LLM

© 2024 IBM Corporation



Customer starts with:
what is ITSM?

Knowledge: Understanding IT Service Management (ITSM) IT Service Management (ITSM) refers to the entirety of...

Prompt Instructions: Default

granite-13b-chat-v2

Understanding IT Service Management (ITSM)
IT Service Management (ITSM) refers to the entirety of activities, policies, processes, and procedures that organizations use to design, deliver, manage, and improve the way IT services are used within their business. ITSM ensures that the right mix of people, processes, and technology are in place to provide value to customers.
Key Components of ITSM.
1. Incident Management: Restoring normal service operation as quickly as possible.
2. Problem Management: Identifying the root cause of incidents and nding permanent solutions.
3. Change Management: Managing changes to IT services with minimal disruption.

Add prompt instructions

A default prompt is provided to produce the best general responses. Optionally, you may add instructions to refine the responses from the LLM. These instructions complement the default prompt, but are not integrated as part of it.

This final step concludes the lab. You just configured and tested conversational search to provide content-grounded answers to users' questions. You also experimented with general-purpose answering using LLMs. Finally, you created an AI-guided action to answer questions from a user related to a specific subject, given provided knowledge.

© 2024 IBM Corporation