Hands-on Lab: Explore IBM watsonx.ai

Estimated time needed: 30 minutes

Welcome to the hands-on lab explore IBM watsonx.ai!

What is IBM watsonx.ai?

Watsonx is IBM's enterprise-ready AI and data platform that is designed to multiply the impact of AI throughout your organization. The platform includes three products: watsonx.ai studio, watsonx.data *fit-for-purpose data store*, and watsonx.governance toolkit.

This lab focuses on watsonx.ai studio for developing new foundation models, generative AI, and machine learning. It combines new generative AI capabilities, powered by foundation models, and traditional machine learning into a powerful studio spanning the AI lifecycle. With IBM watsonx.ai, you can easily train, validate, tune, and deploy generative AI foundation models and machine learning capabilities. You can also build AI applications in a fraction of the time with a fraction of the data.

In this lab, you will explore the Prompt Lab of IBM watsonx.ai, using various prompts to experiment with foundation models. You will experiment with various natural language processing (NLP) tasks, including question-answering, content generation, text summarization, text classification, and text extraction.

Learning objectives

After completing this lab, you will be able to:

- Explore generative AI capabilities of IBM watsonx.ai
- Explore IBM watsonx.ai Prompt Lab and use prompts to experiment with foundation models
- Use Prompt Lab of IBM watsonx.ai for a range of natural language processing tasks to generate content

Prerequisites

You will need an IBM Cloud account to sign up for watsonx.ai. Please ensure you have completed the previous lab (Creating an IBM Account) for creating an IBM Cloud account before starting this lab.

Exercise 1: Get started with watsonx.ai

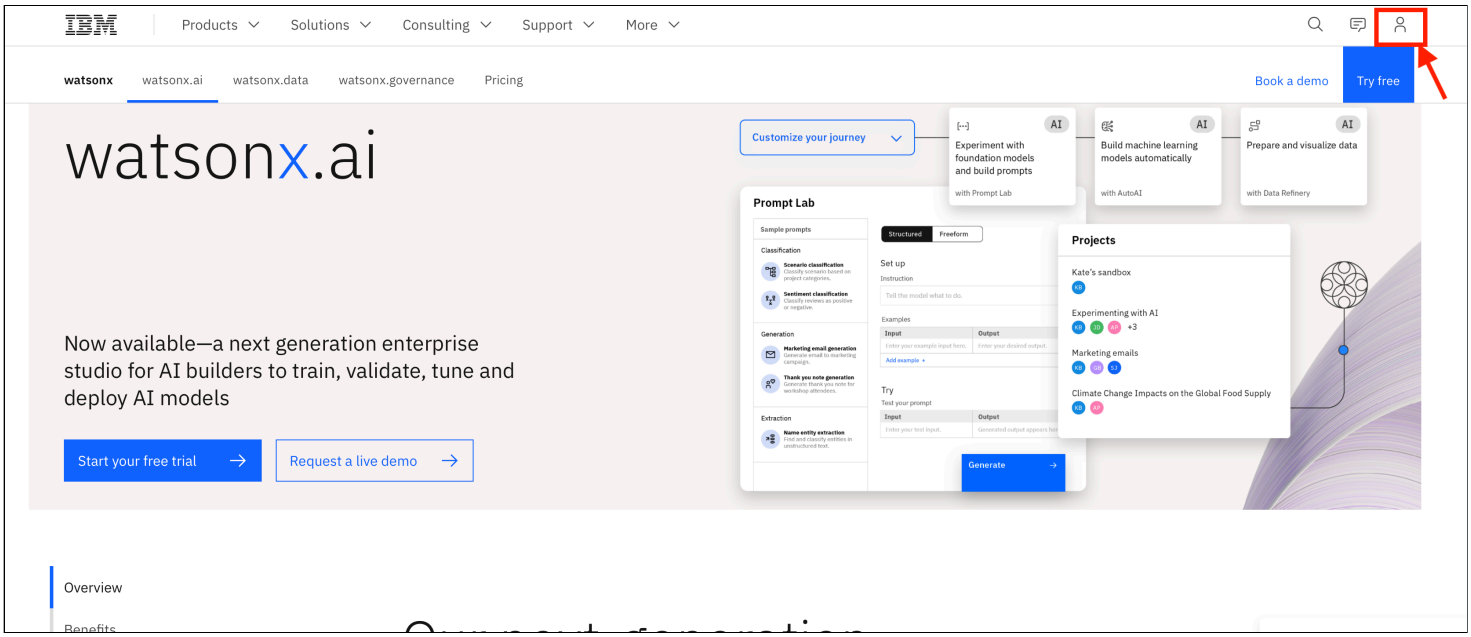
In this exercise, you will generate content using IBM's watsonx.ai Prompt Lab. You will experiment with foundation models in the Prompt Lab using various prompts without any coding requirements.

> Note: The Prompt Lab is a foundation model tool, not a chatbot interface.

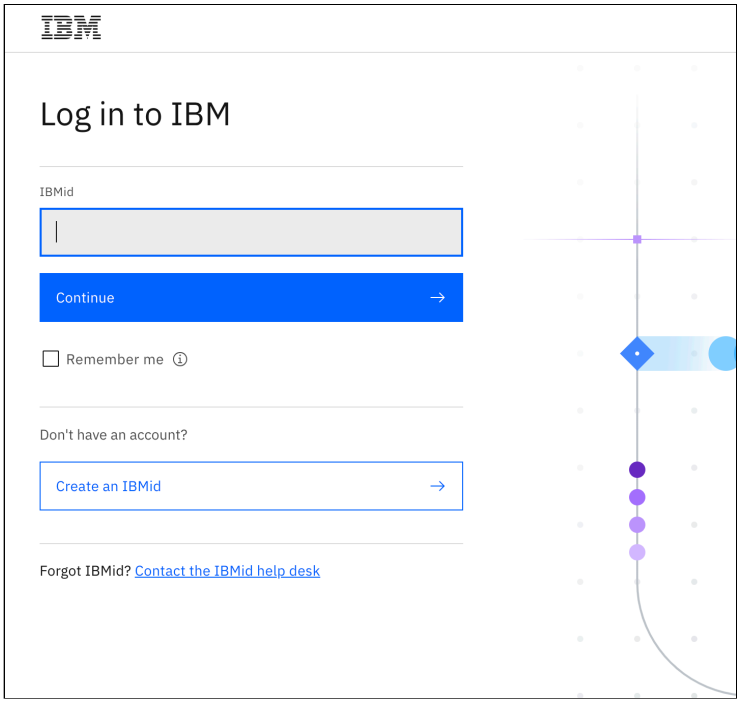
Steps:

1. Click IBM watsonx to launch it.

2. Click the user icon on the upper right corner of the IBM watsonx.ai home screen to log in.

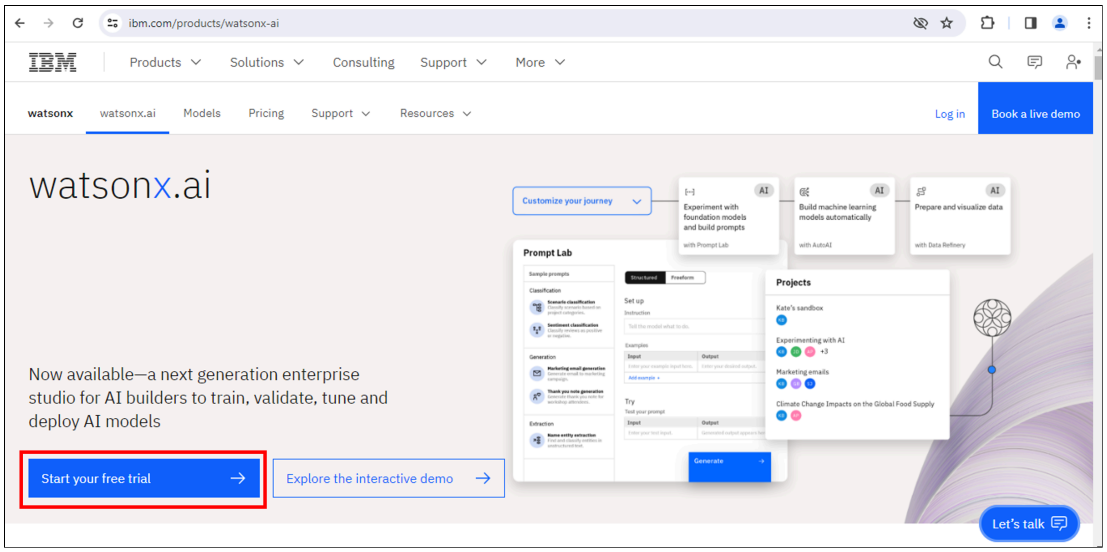


3. Log in using the credentials you used while signing up with the feature code.

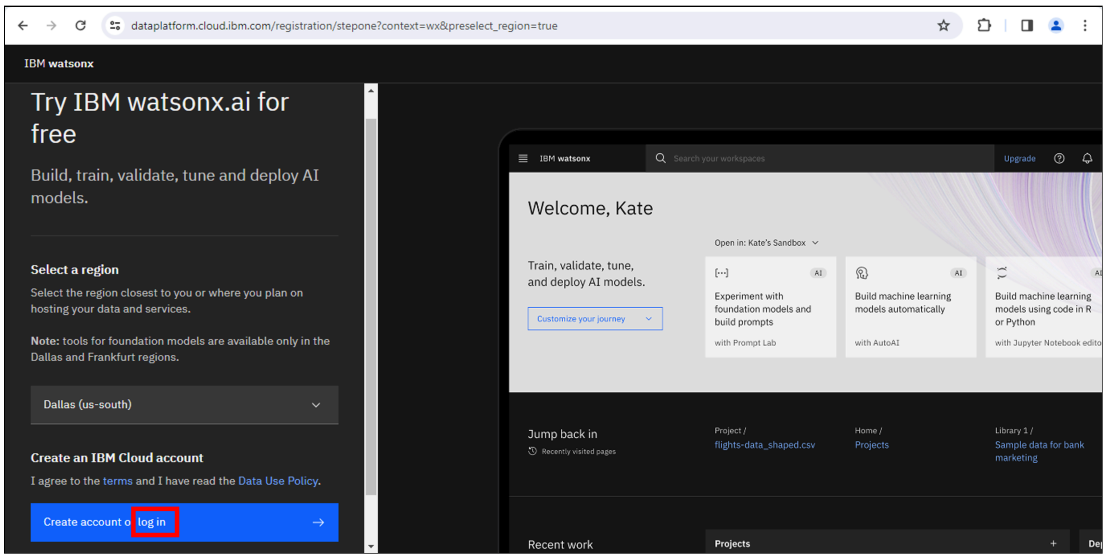


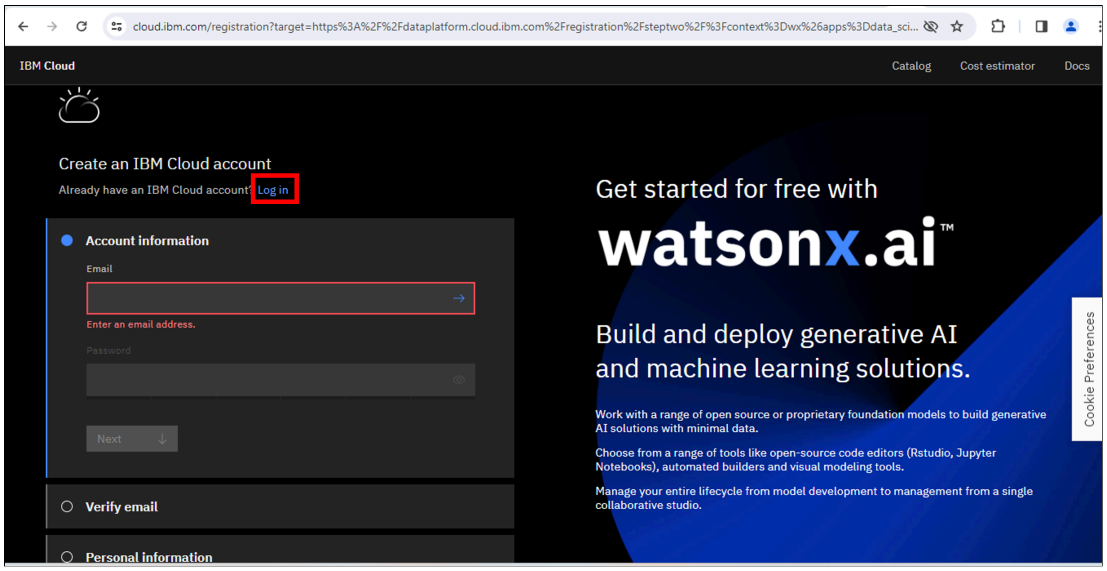
Note: You might receive a verification code via email to verify and login to your IBM watsonx account.

4. After logging in, you will land on the home page, where you need to click 'Select your free trial' and log in again.

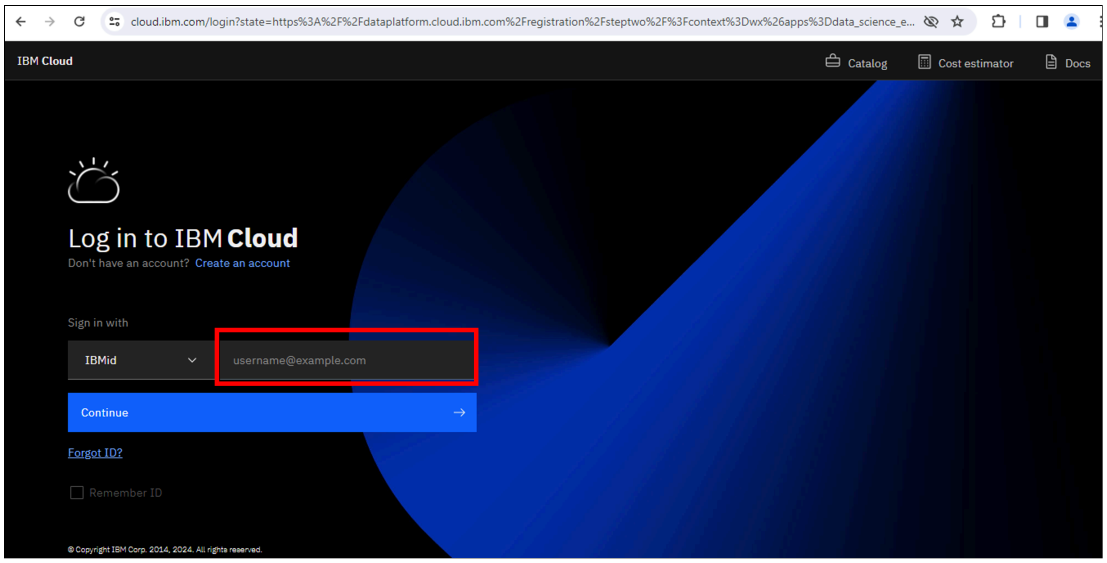


5. Click on **Create account or Log in**. In the next step, click on **Log in**, along the text Already have an IBM Cloud account.

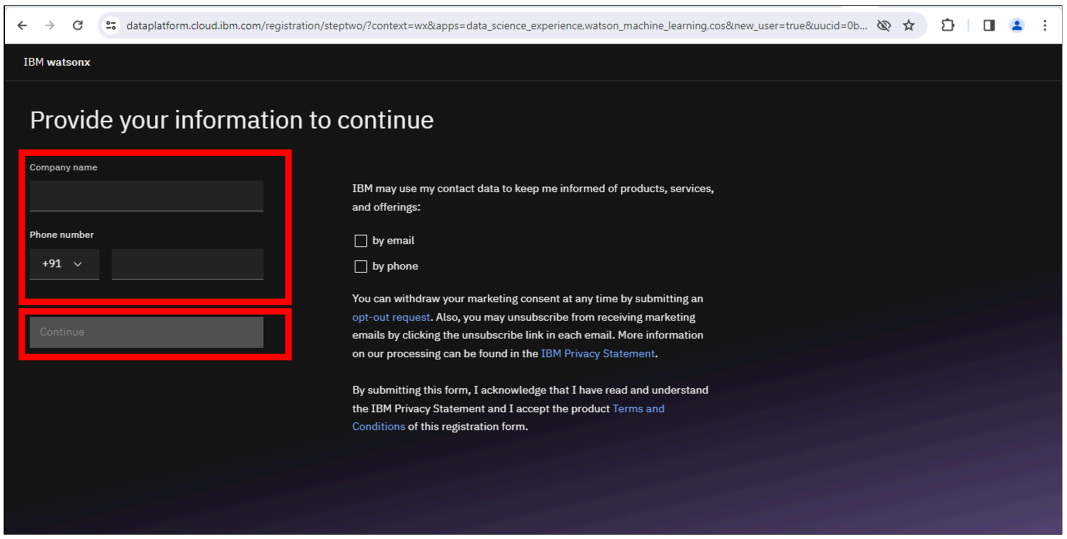




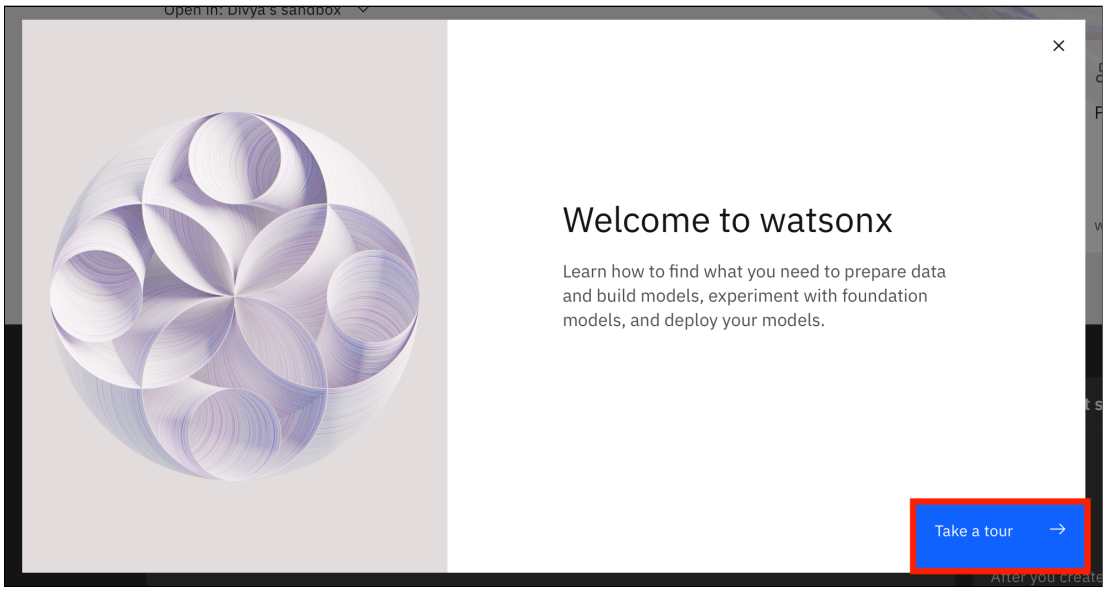
6. Log in again using the credentials you used while signing up with the feature code.



7. Provide the required information and then click on **Continue**.

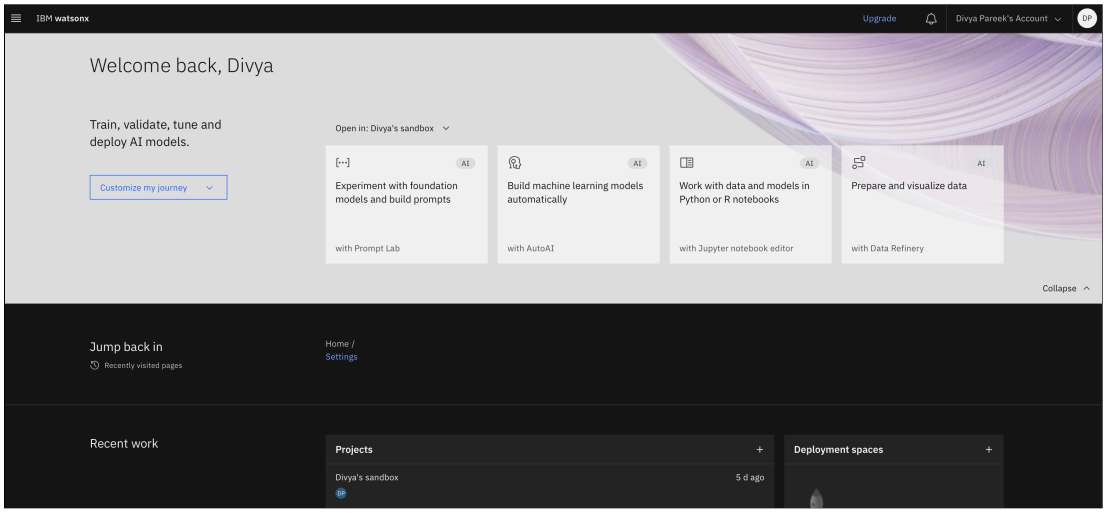


8. After the successful login, you will be prompted to **Take a tour** of the watsonx platform (on your first login), as shown below.



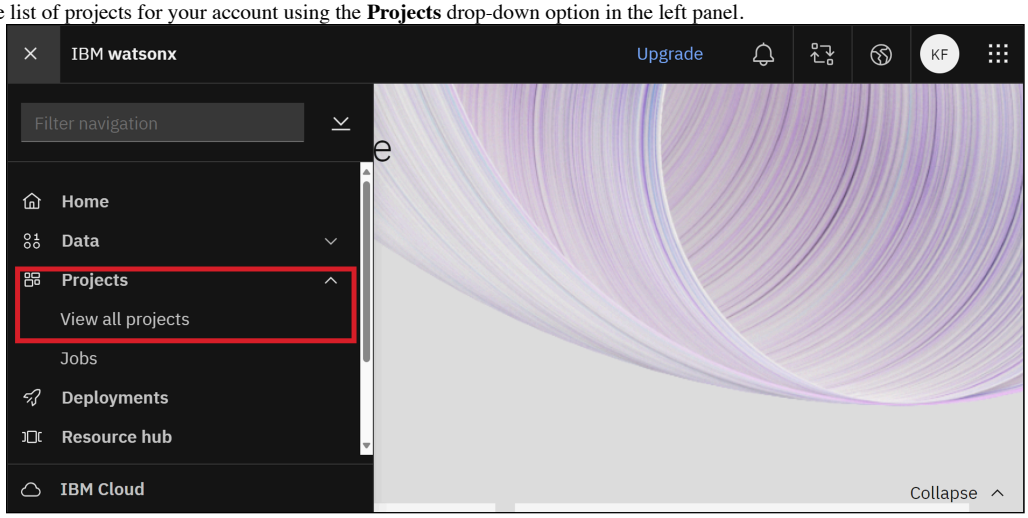
Note: You may opt for the watsonx walkthrough by clicking **Take a tour** or skip or exit it by clicking **X** at the upper right corner of the screen

9. On completing or exiting the tour, you will reach your IBM watsonx.ai dashboard page, as shown below. Here, you can perform experiments like train, validate, tune, and deploy AI models.

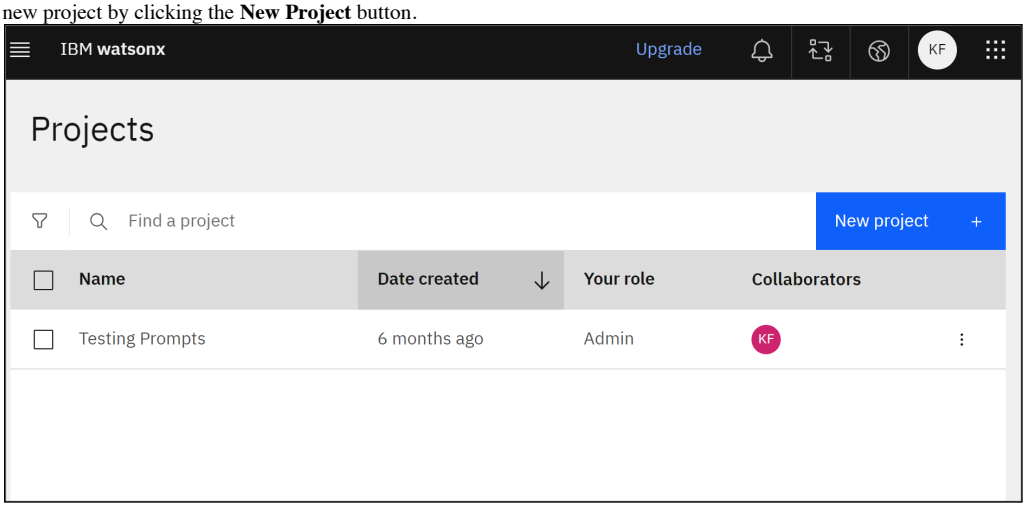


Note: If this is NOT your first login session in the watsonx.ai, you may find the **Prompt Lab** deactivated. In this case, you need to create a project and then select the **Experiment with foundation models amd build prompts** option. You can create a new project by following the given steps.

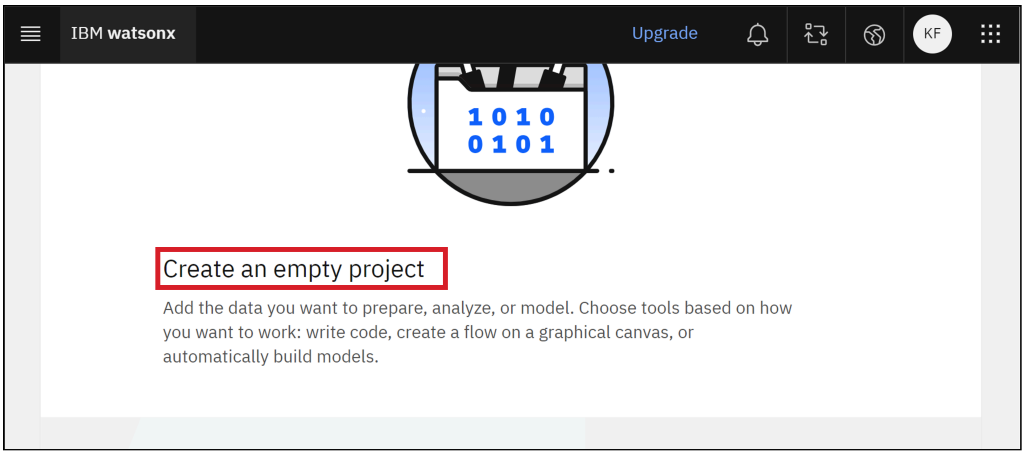
a. You can view the list of projects for your account using the **Projects** drop-down option in the left panel.



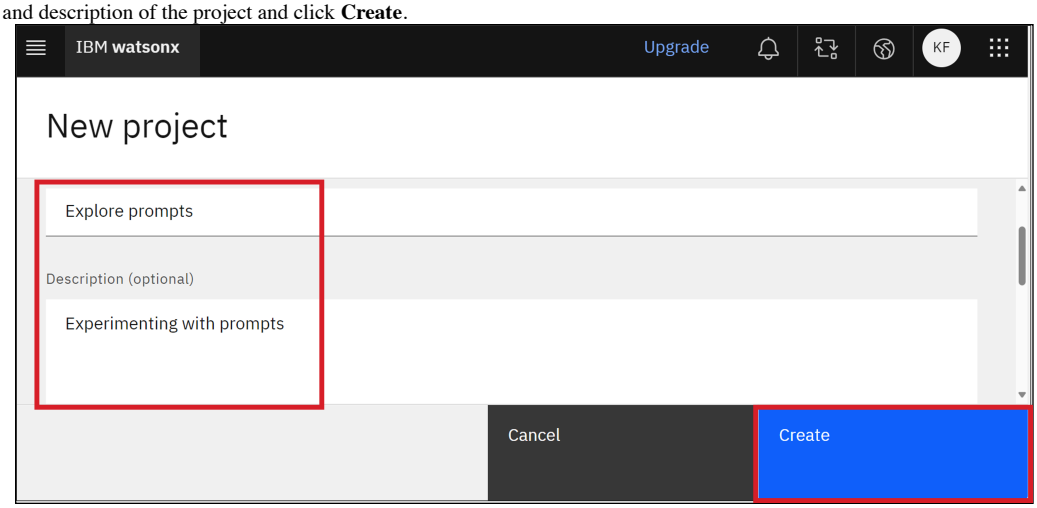
b. You can create a new project by clicking the **New Project** button.



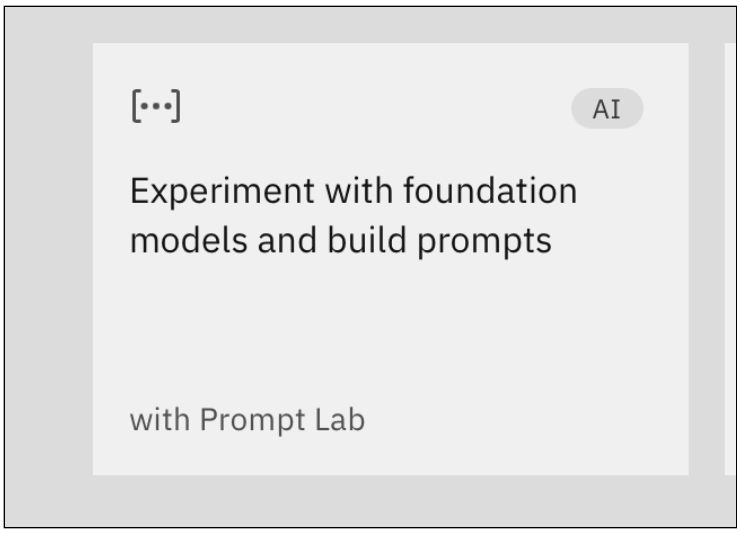
c. Select the option for the empty project.



d. Enter the name and description of the project and click **Create**.

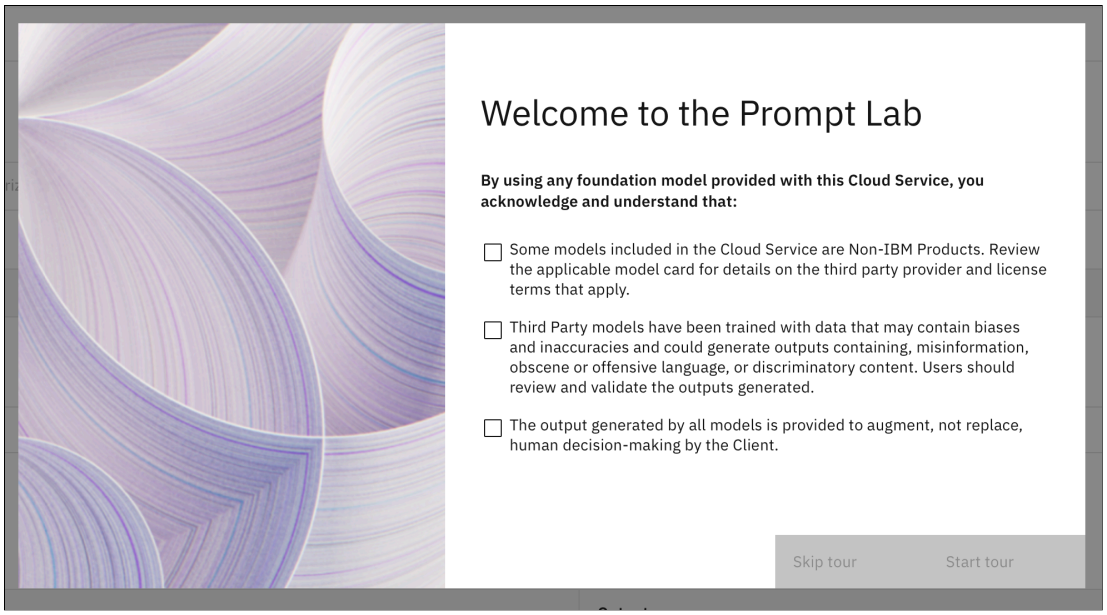


10. To experiment with a language model, click the first card **with Prompt Lab**, as shown in the screenshot below.

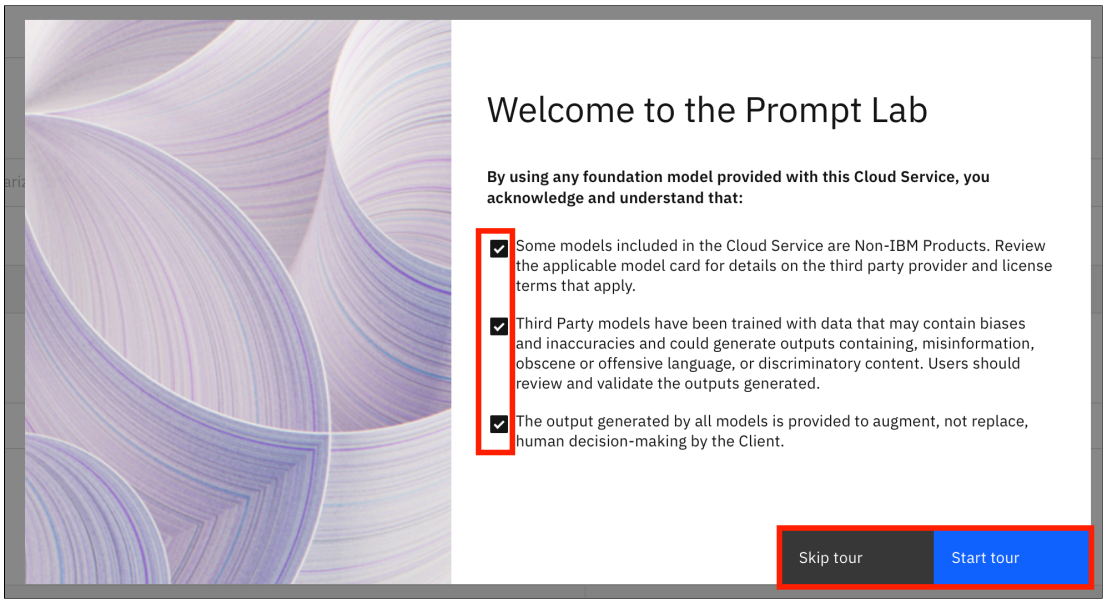


Note: If you have created a new project, click **Experiment with foundation models and build prompts** in the Start working section

11. You will view the **Welcome to the Prompt Lab** tour page, as shown below.



12. You must check all three acknowledgment boxes to start or skip the tour, as shown below.

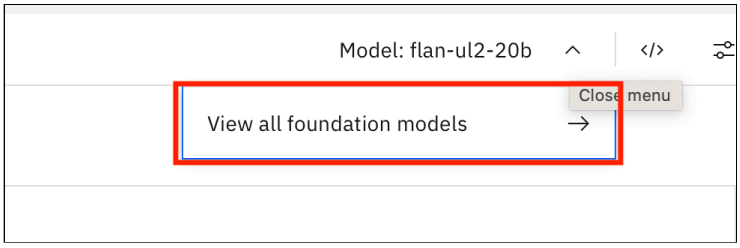


13. You can use the foundation model **flan-ul2-20b** from the upper right corner of the sandbox area.

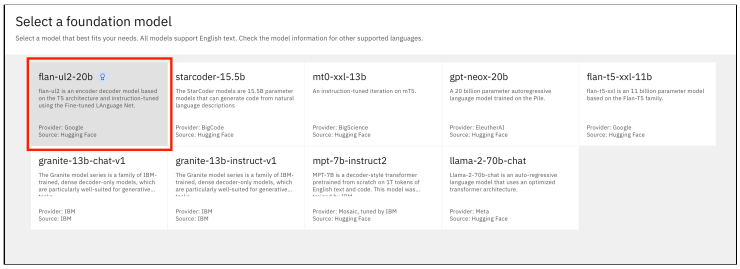
Note: The flan-ul2-20b model is an encoder-decoder model based on the T5 architecture. This model may appear as the default selection. Alternatively, you can select it from the list.

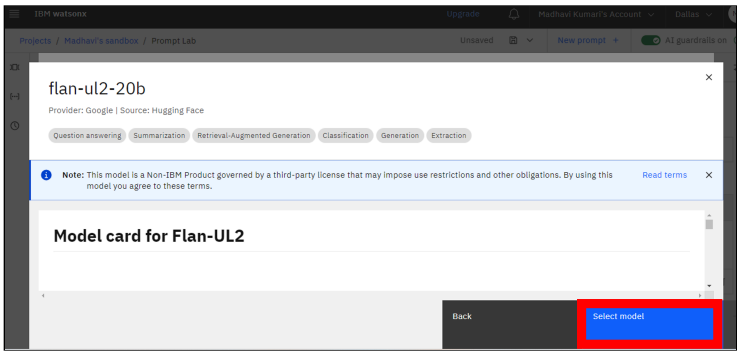
a. Select the drop-down arrow next to the model name.

b. Select **View all foundation models** from the drop-down option.



c. You will view various foundation models. Select **flan-ul2-20b** model from the available foundation model options, as shown below.





The foundation model **flan-ul2-20b** model can help you generate text such as poems, code, and scripts. It can assist you in translating languages, creating creative content such as email letters, and providing informative answers to your questions.

When generating text or answering questions, Google AI's large language model, a 20 billion parameter model trained on a massive data set of text and code, can consider up to 4096 tokens of context.

14. Once the foundation model is selected, you will reach the sandbox workspace area. You will see the **Instruction (optional)** text box below **Set up** under the **Structured** option of the foundation model.

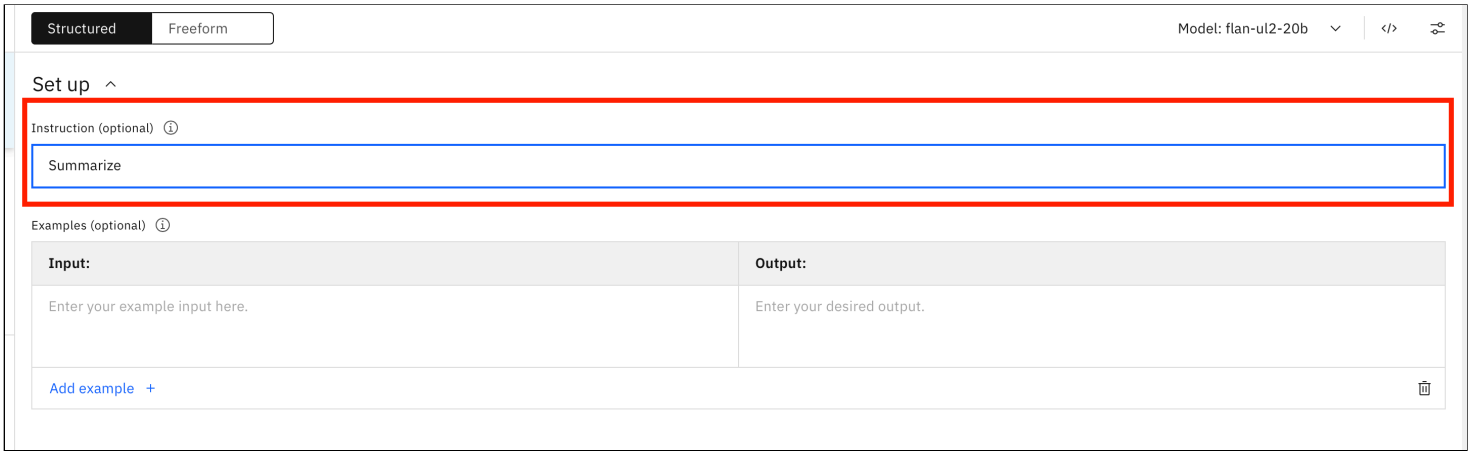
Note: To perform any NLP–based task, you can enter the prompt type of your choice in the **Instruction (optional)** text box.

15. Think of an NLP-based task and enter it in the **Instruction (optional)** text box, as shown below.
    For example, for a text summarization task, you can instruct the model as:

```
1. 1
```

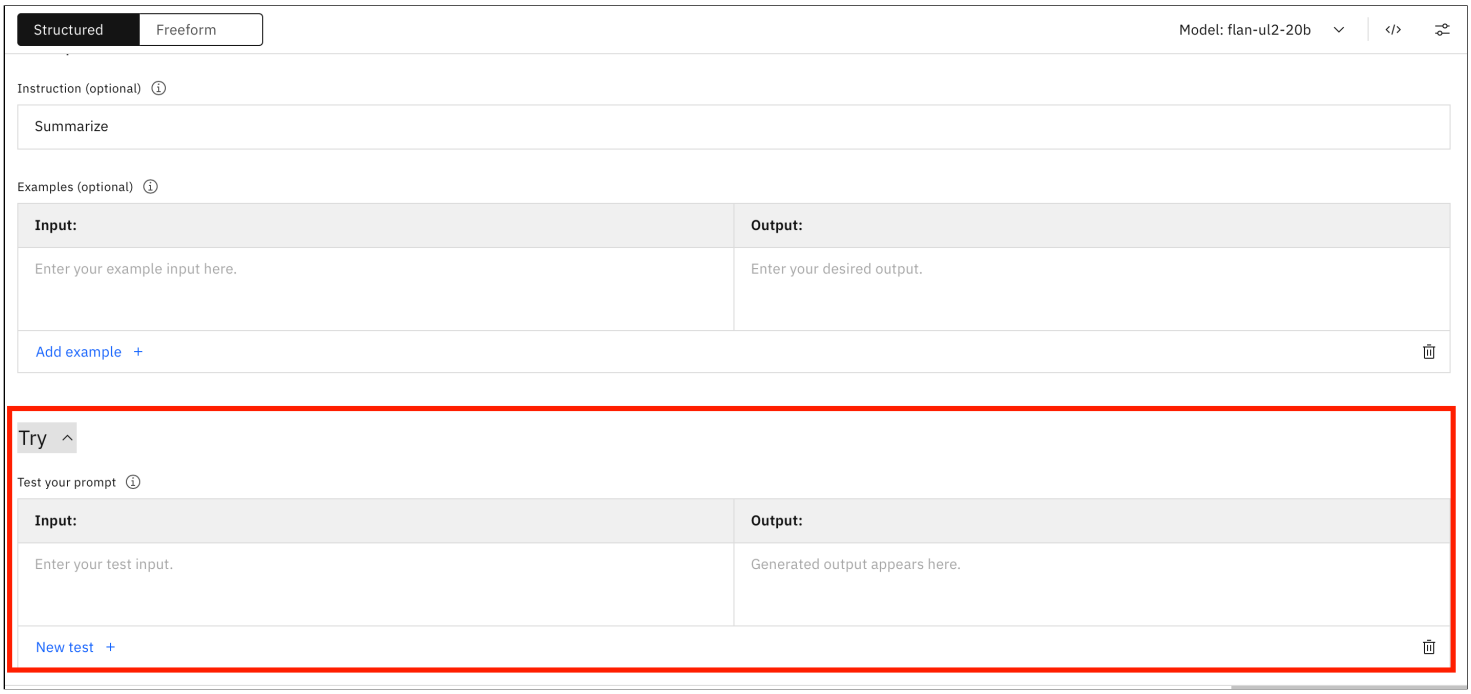
```
1. Summarize
```

Copied!



Note: You can skip the **Example (optional)** text box.

16. You will see the **Try** section on the bottom part of the screen, as shown below.



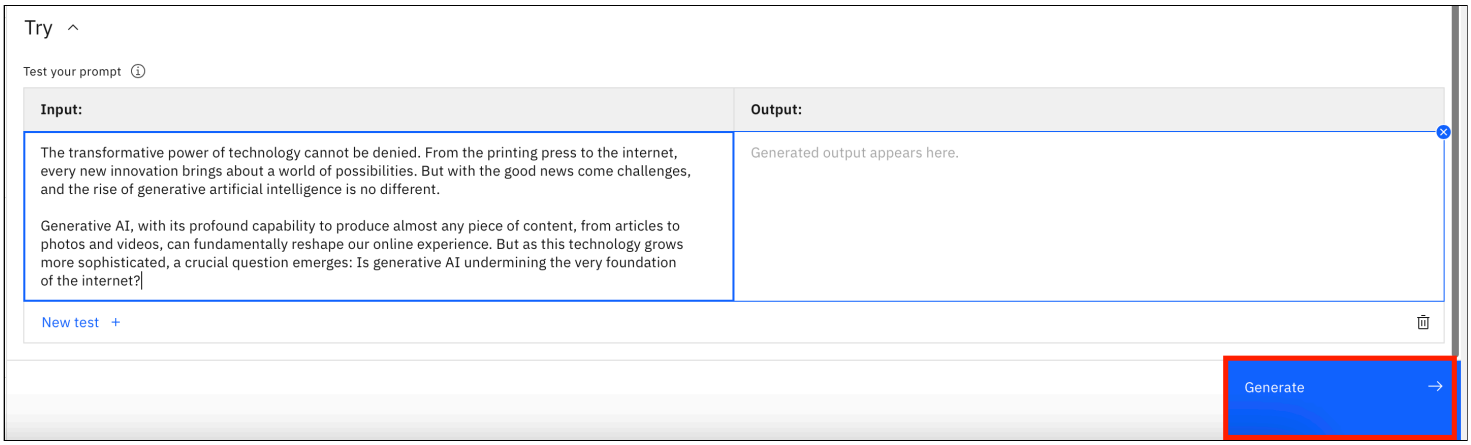
17. In the **Input** text box of the **Try** section, enter the text you wish to summarize and click **Generate**.
    For example,

```
1. 1
2. 2
```

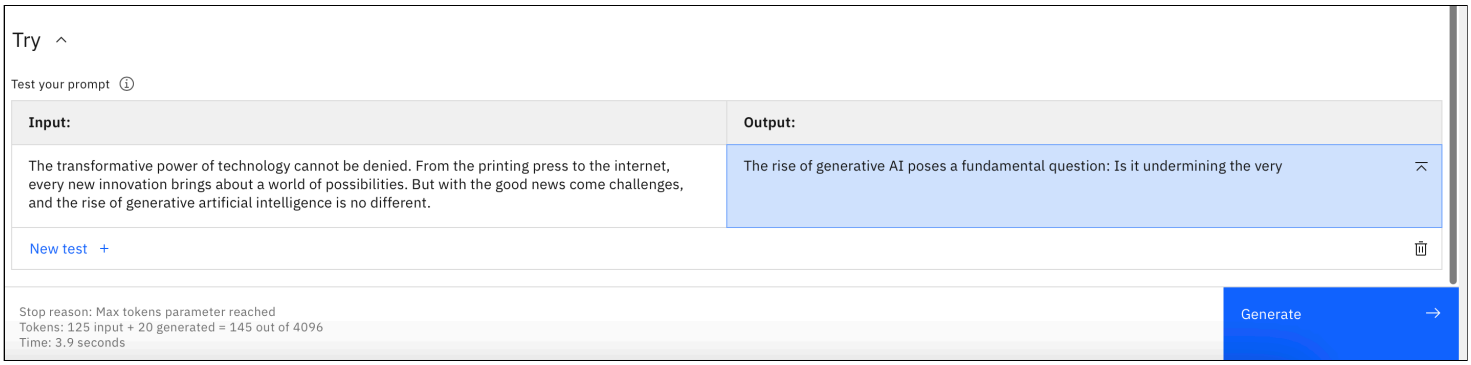
```
1. AI’s value is not limited to advances in industry and consumer products alone. When implemented in a responsible way—where the technolog
2. Despite the many potential advantages, many government agencies are still grasping on how to implement AI and generative AI in particula
```

Copied!

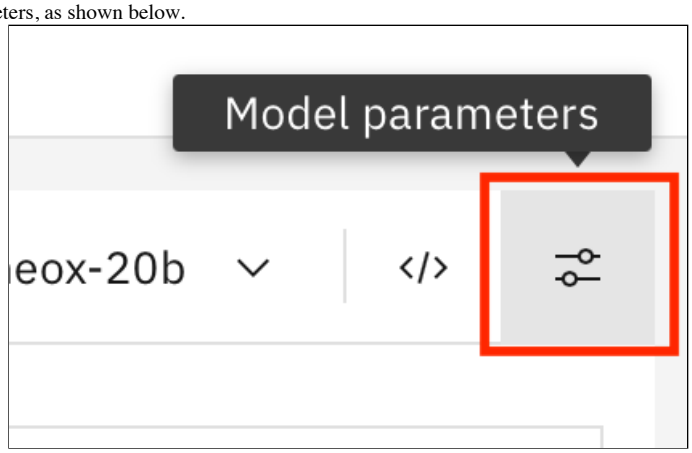
Source: [IBM blog: What can AI and generative AI do for governments?]



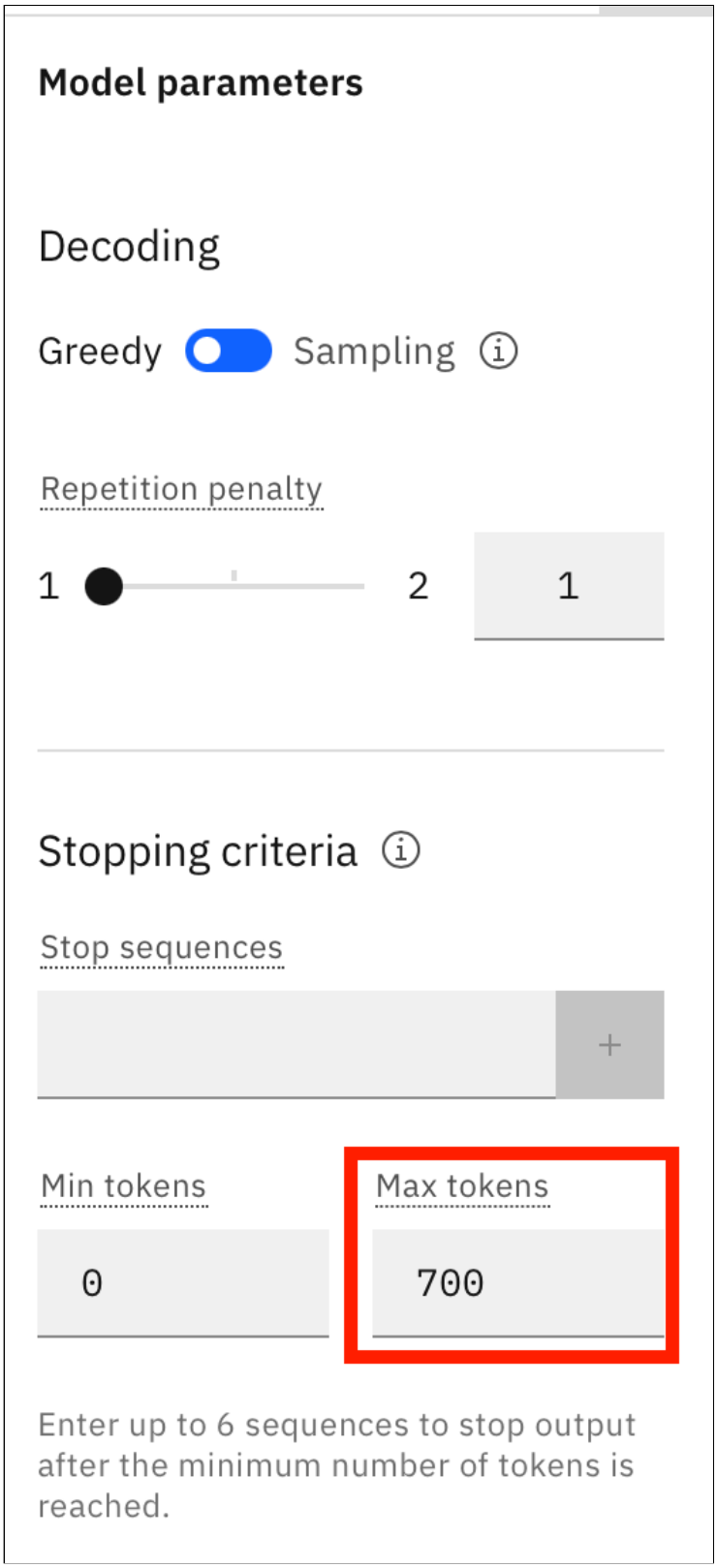
18. Text summary will appear in the **Output** box of the **Try** section.



Note: If you get an incomplete answer, you must check and update the output token limit in the model parameters section. Go to the upper right corner of the workspace and click the model parameters, as shown below.



In the **Stopping criteria** section, you can enter the output token limit up to 700 tokens for this model, as shown below.

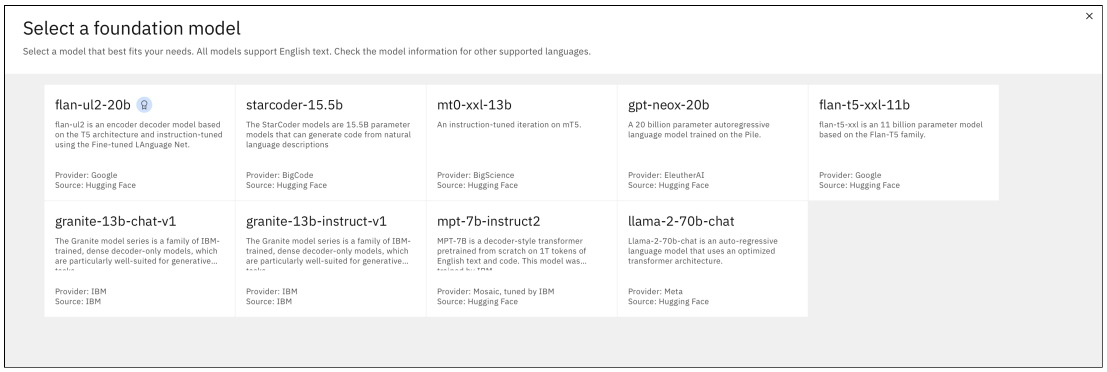


Note: The output token limit may vary from model-to-model, so you must pay attention to the pop-up notifications and reset the token limit accordingly. The highest number of tokens can be set to 1024 tokens.

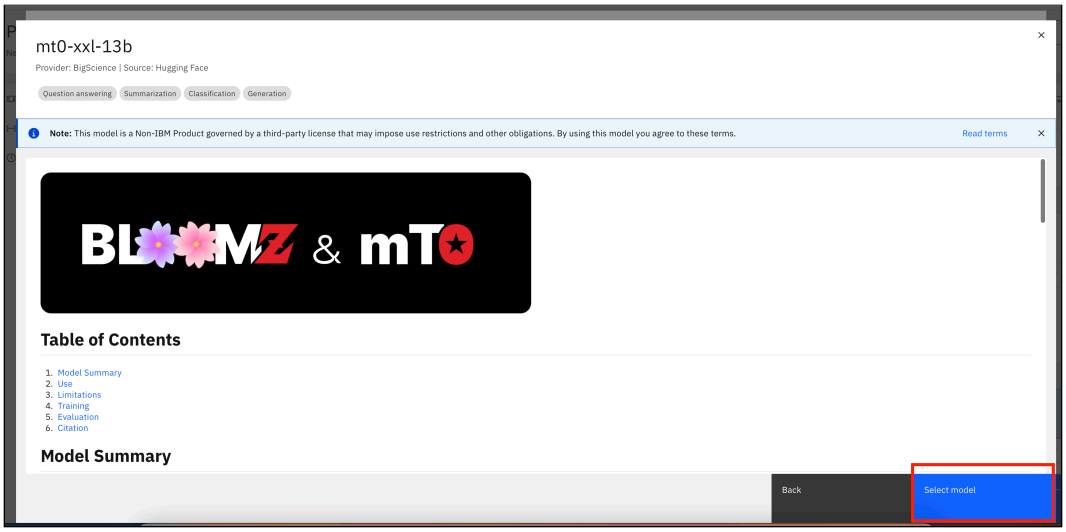
Exercise 2: Explore various foundation models in watsonx.ai

IBM watsonx.ai includes various foundation models in this Prompt Lab sandbox environment. You must experiment with different foundation models in this Prompt Lab and explore their intended use and capabilities.

1. Select a foundation models from the upper right corner of the sandbox environment, as shown below.



2. Review the details of the chosen foundation models in the pop-up window, as shown below, and click **Select model** to select this model for prompt testing.



3. The selected model will be reflected in the upper right corner, as shown below.

4. Provide relevant prompts to get the desired output from the selected model.

5. Check the output quality by submitting alternatives to make your prompts more precise.

Let's learn about some of these foundation models and their intended use and capabilities.

1. Foundation model mt0-xxl-13b

The foundation model **mt0-xxl-13b** is a 13-billion-parameter model trained on a massive text and code data set in 40 languages. It can summarize factual topics, provide informative answers to your questions, and translate languages into multiple languages.

It can be used to perform tasks expressed in natural languages, such as:

1. `Search related terms`
2. `Write text or stories`
3. `Translate languages`
4. `Answer queries in various languages`

Try the following exercises using the **mt0-xxl-13b** foundation model:
`1. Provide a term and ask the model to suggest related search terms.`
For example,
Instruction:

```
1. 1
```

```
1. Search related terms
```

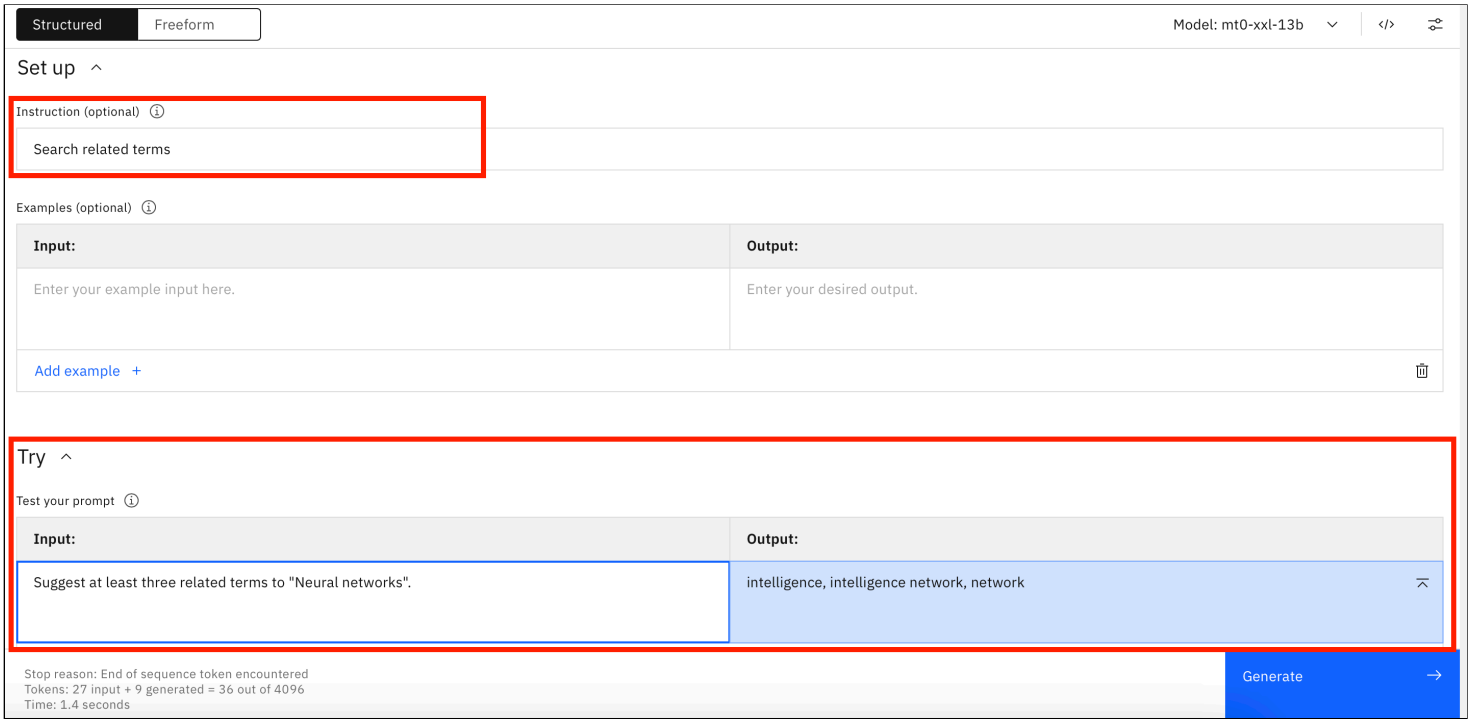
Copied!

Prompt:

```
1. 1
1. Suggest at least three related terms to "Neural networks
```

Copied!

You will see a response similar to the screenshot below.




```
2. Give a question and seek the answer in another language.
```
For example,
Instruction:

```
1. 1
1. Answer in Spanish
```

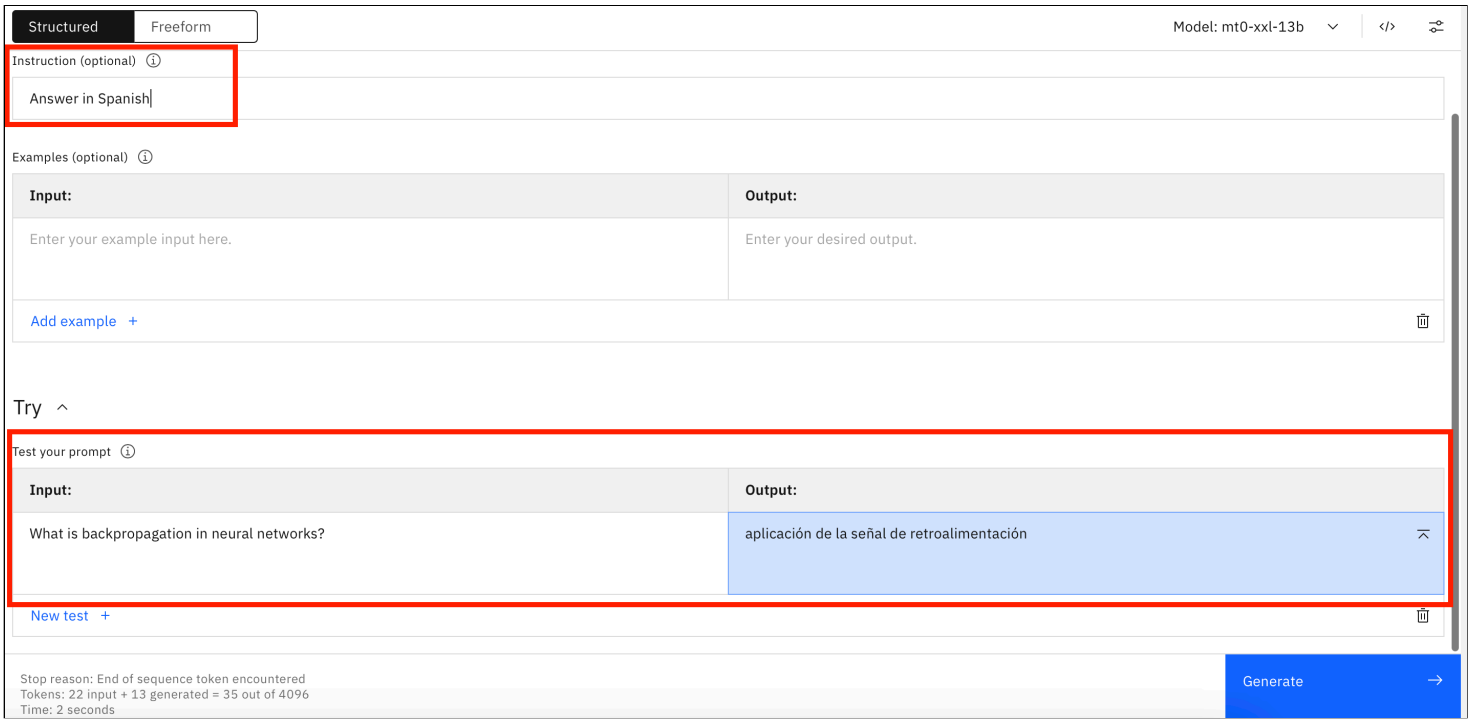
Copied!

Prompt:

```
1. 1
1. What is backpropagation in neural networks?
```

Copied!

You will see a response similar to the screenshot below.



Limitations of the model: It is open source, so anyone can use or modify it. It is still in development; hence, it may not be able to handle all types of tasks perfectly.

2. Foundation model flan-t5-xxl-11b

The foundation model **flan-t5-xxl-11b** is a large language model chatbot created by Google AI. It is a 340-billion-parameter model that was trained on a massive data set of text and code to perform various natural language tasks.

It can be used to perform tasks expressed in natural languages, such as:

1. `Text completion`
2. `Answer questions`
3. `Text summarization`
4. `Language translation`

Try the following exercise using the **flan-t5-xxl-11b** foundation model:
`1. Get the answer and explanation to a question.`
For example,
Instruction:

```
1. 1
```

```
1. Answer the question in English
```

Copied!

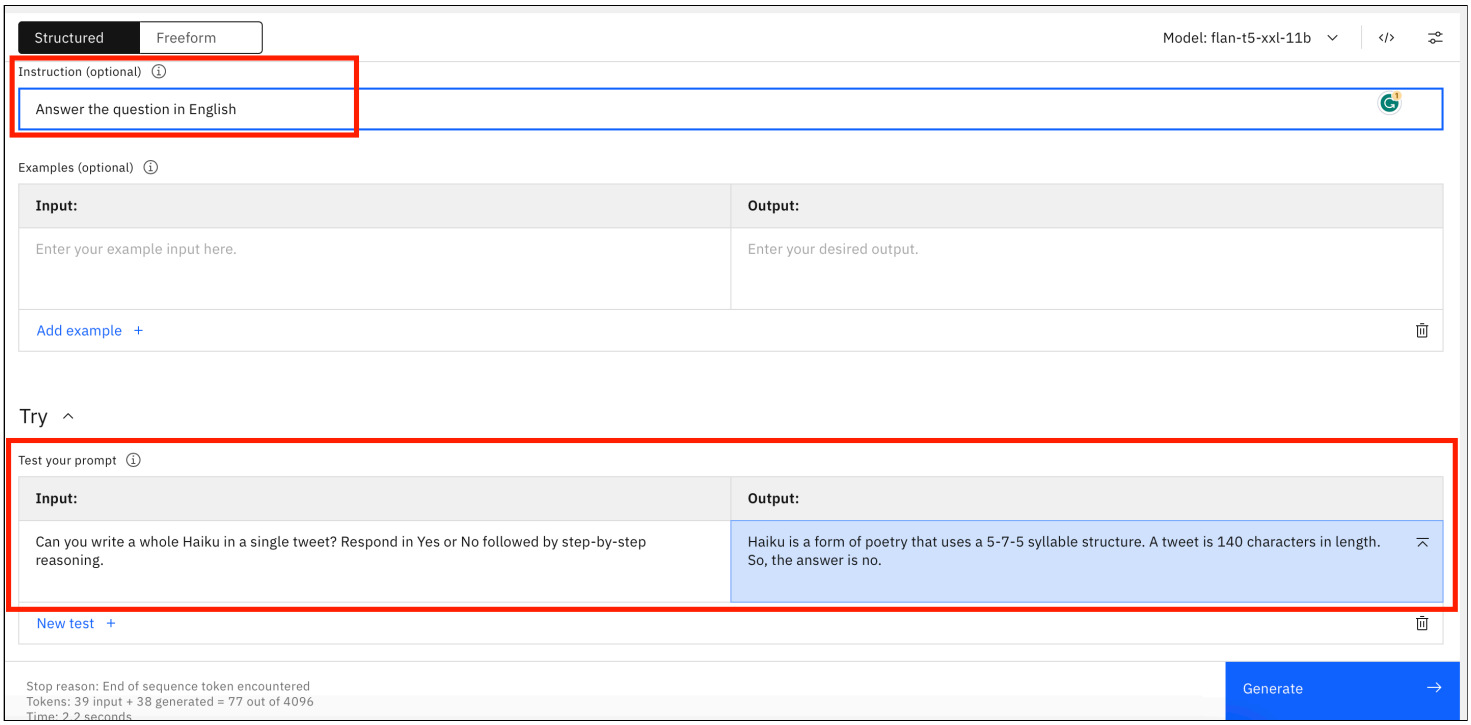
Prompt input:

```
1. 1
```

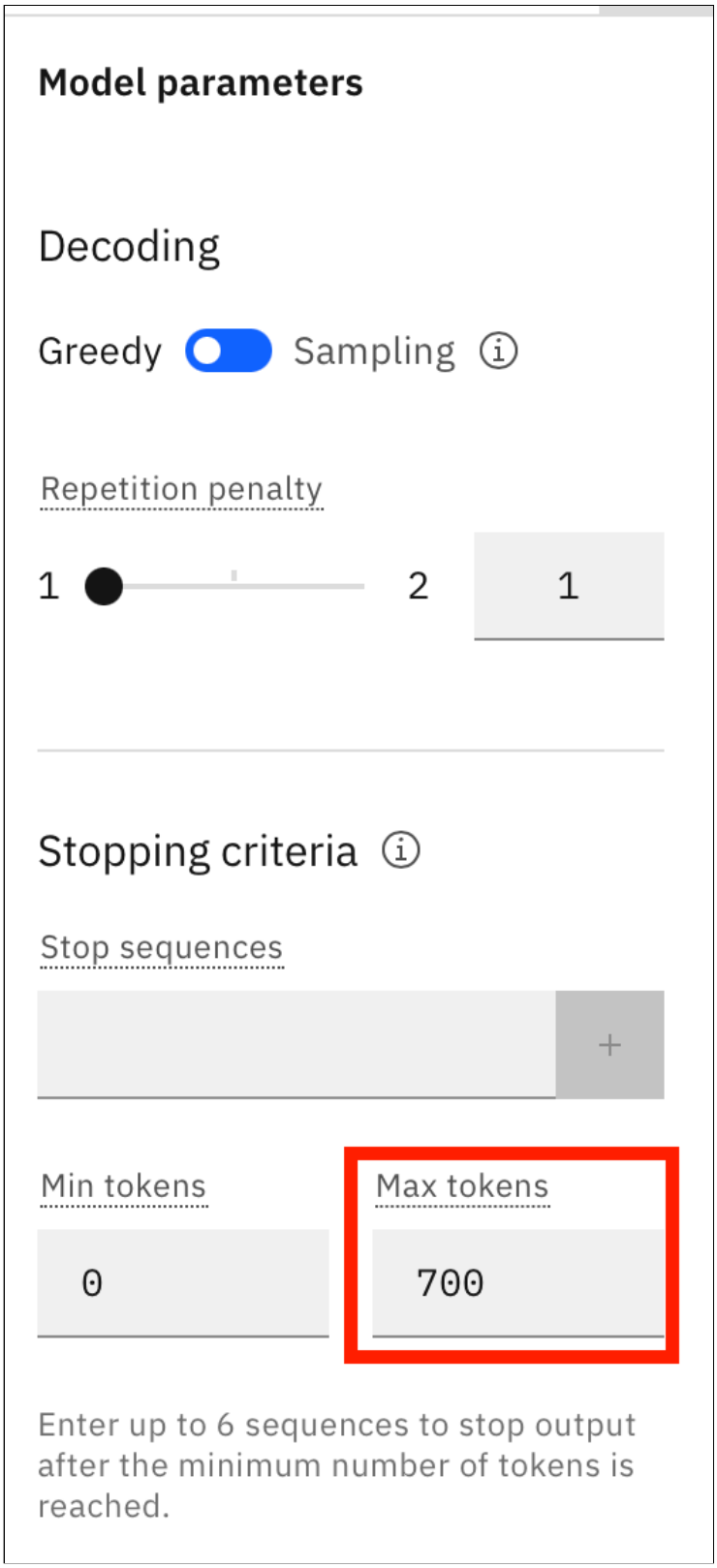
```
1. Can you write a whole Haiku in a single tweet? Respond Yes or No followed by step-by-step reasoning.
```

Copied!

You will see a response similar to the screenshot below.



Note: If you get an incomplete answer, you must check and update the output token limit in the model parameters section. Go to the upper right corner of the workspace, click the model parameters, and set the maximum tokens to 700, as done in step 1.



Limitations of the model: It may carry data bias and, hence, may generate unreliable output. It requires huge power and memory resources to run, which makes it difficult to apply on a smaller scale.

3. Foundation model gpt-neox-20b

The foundation model **gpt-neox-20b** is a 20 billion parameter autoregressive language model developed by EleutherAI. It is trained on the Pile using the GPT-NeoX library. It was developed primarily for research purposes. This model works with the Transformers library.

It can be used to perform the following natural language tasks:

1. `Generate text`
2. `Extract information`

3. `Summarize the text`

Try the following exercise using the **gpt-neox-20b** foundation model:
`1. Provide an email text and extract useful information from it, such as mailing address.`
For example,
Instruction:

```
1. 1

1. Extract the information
```

Copied!

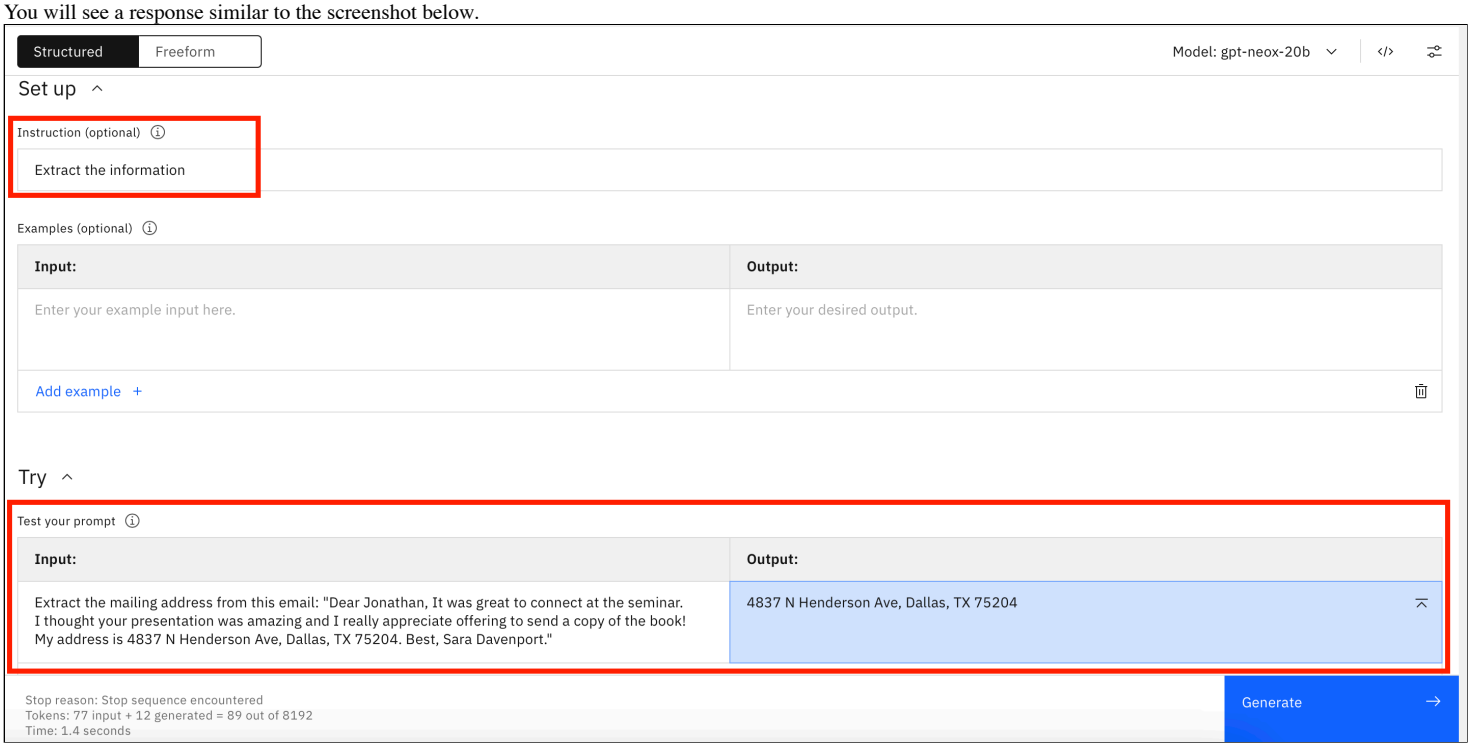
Prompt:

```
1. 1

1. Extract the mailing address from this email: "Dear Jonathan, It was great to connect at the seminar. I thought your presentation was ama
```

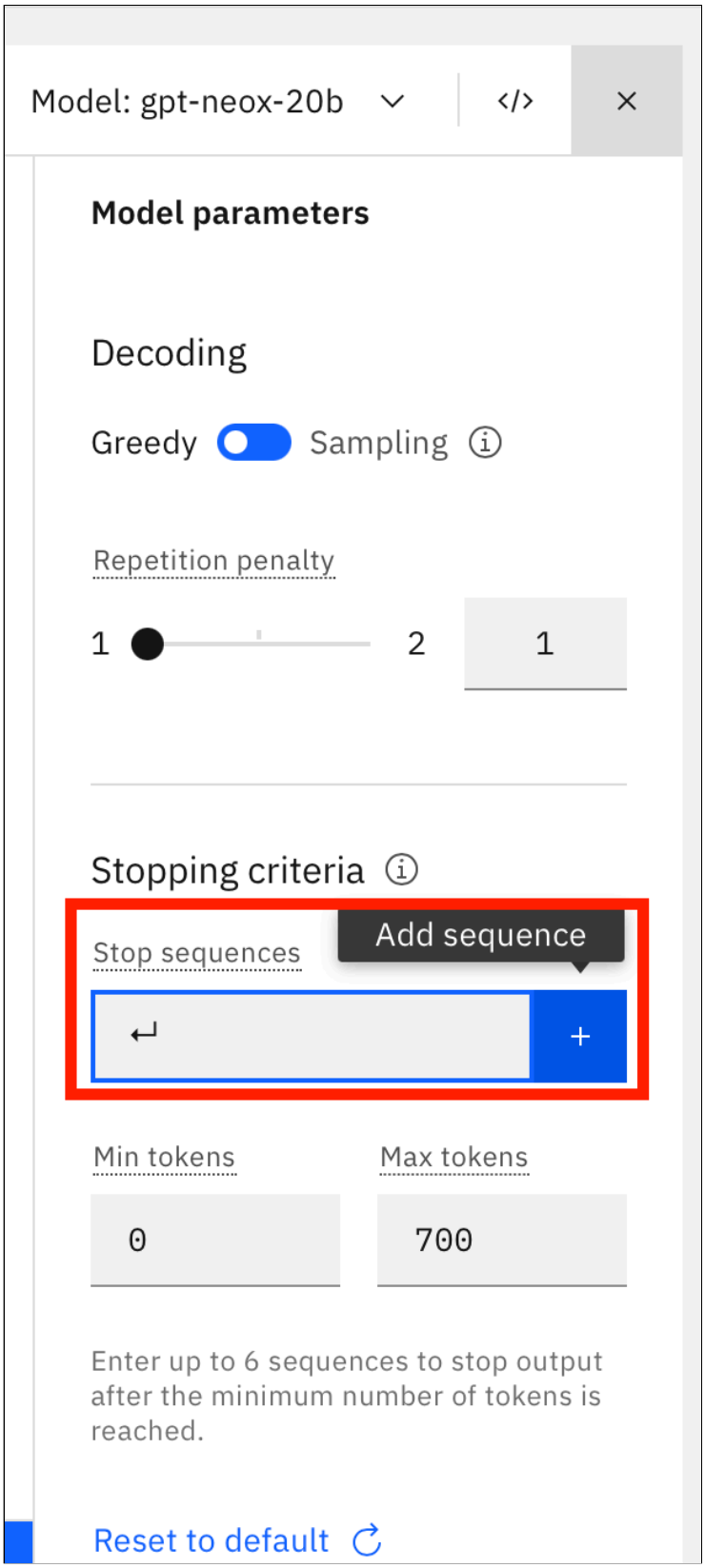
Copied!

You will see a response similar to the screenshot below.



Note: You might experience that the output is repeating the information.
To fix this, you need to go to model parameters from the upper right corner of the workspace, as shown below.

Now, press the "enter" key in the **Stop sequences** box given under Stopping criteria, and click the "+" (plus sign) to add it to the stop sequences criteria.



Limitations of the model: It is an English-language model that cannot be used for a variety of downstream tasks in other languages. To fine-tune your application model using this foundation model, you must conduct your risk and bias assessment.

4. Foundation model mpt-7b-instruct2

The foundation model **mpt-7b-instruct2** is a large language model chatbot developed by IBM. It is a 7 billion parameter model that is trained on a massive data set of text and code. It was developed primarily for research purposes. This model works with the transformers library. It can provide informative answers to your questions, even if they are open-ended, difficult, or strange. It can generate various creative text formats, such as poems, code, scripts, musical pieces, email, letters, and so on.

It can be used to perform natural language tasks, such as:

```
1. Text generation
2. Language translation
3. Creative content creation
4. Informative answers to questions
```

Try the following exercise using the **mpt-7b-instruct** foundation model:

```
1. Provide a long conversation between two or more parties and ask the model to write a summary of the discussion given in the text.
```

For example,
Instruction:

```
1. 1

1. Summarize the text
```

Copied!

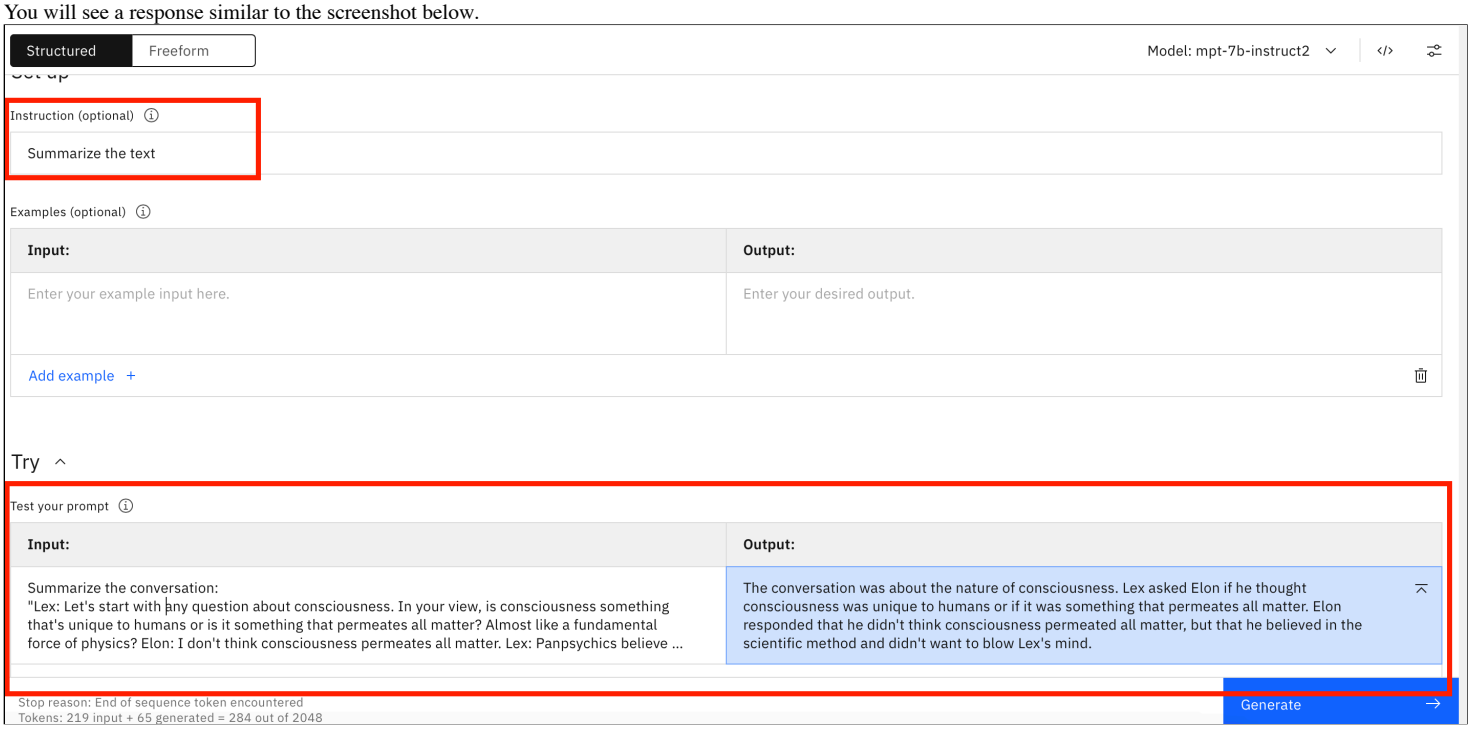
Prompt:

```
1. 1

1. Summarize the conversation: "Lex: Let's start with an easy question about consciousness. In your view, is consciousness something that's
```

Copied!

You will see a response similar to the screenshot below.



> Note: You may check the token limit if the generated response seems incomplete.

Limitations of the model: It is still in development, so it may not be able to handle all types of tasks perfectly. It can be computationally expensive to train and use.

> Note: This lab covers some of the foundation models available, as the watsonx.ai environment is continuously evolving and expanding to more foundation models. You may explore and experiment with more foundation models in the watsonx.ai environment.

Summary:

Congratulations! You just completed the hands-on lab explore IBM watsonx.ai.

In this lab, you learned about the generative AI foundation models of IBM watsonx.ai. You used various prompts to experiment with various foundation models using Prompt Lab. You also generated content for various natural language processing (NLP) tasks, including question-answering, content generation, text summarization, text classification, and text extraction.

© IBM Corporation 2023. All rights reserved.