

# **Module 1**

## **Reading 1: Course Overview**

This course is designed for practitioners and enthusiasts passionate about learning prompt engineering techniques to unlock the full potential of generative artificial intelligence (AI) models.

This course explains the techniques, approaches, and best practices for writing effective prompts. You will learn to utilize these techniques to effectively guide generative AI models and control their output to obtain the desired results.

You will learn about prompt techniques like zero-shot and few-shot, which can improve the reliability and quality of large language models (LLMs). You will also explore various prompt engineering approaches like Interview Pattern, Chain-of-Thought, and Tree-of-Thought, which aim at generating precise and relevant responses.

You will be introduced to commonly used prompt engineering tools like IBM watsonx Prompt Lab, Spellbook, and Dust.

The hands-on labs included in the course offer an opportunity to optimize results by creating effective prompts in the IBM Generative AI Classroom. You will also hear from practitioners about the tools and approaches used in prompt engineering and the art of writing effective prompts.

After completing this course, you will be able to:

- Explain the concept and relevance of prompt engineering in Generative AI models
- Assess commonly used tools for prompt engineering
- Apply best practices for writing effective prompts
- Apply common prompt engineering techniques and approaches for writing prompts

### **Course Content**

This course is divided into three modules. It is recommended that you complete one module per week or at a pace that suits you - whether its a few hours every day, or completing the entire course over a weekend or even in one day.

#### **Week 1 - Module 1: Prompt Engineering for Generative AI**

In this module, you will learn the concept of prompt engineering in generative AI. You will also learn the best practices for writing effective prompts and assess common prompt engineering tools.

#### **Week 2 - Module 2: Prompt Engineering: Techniques and Approaches**

In this module, you will discover techniques to craft prompts that effectively steer generative AI models. You will also learn about various prompt engineering approaches that can enhance the capabilities of generative AI models to produce precise and relevant responses.

## Week 3 - Module 3: Course Quiz and Wrap-Up

In this module, you will attempt a graded quiz that will test and reinforce your understanding of the concepts covered in the course. The module also includes a glossary to enhance your comprehension of generative AI-related terms. Finally, this module will guide you toward the next steps in your learning journey.

This module also includes optional topics related to text-to-image prompting and prompt engineering capabilities in IBM watsonx.

### Learning Resources

The course offers a variety of learning assets: videos, readings, hands-on labs, expert viewpoints, discussion prompts, and quizzes.

The concepts are presented through videos and readings, complemented by hands-on learning experiences in the labs.

Expert Viewpoints are videos that feature industry practitioners showcasing real-world applications of the skills taught in this course.

Interactive learning is encouraged through **discussions** where you can meet your staff and peers.

**Practice quizzes** at the end of each module will test your understanding of what you learned. The final **graded quiz** will assess your conceptual understanding of the course.

### Who Should Take This Course?

This course caters to anyone eager to delve into the concept of prompt engineering and explore the various tools and approaches.

This course is for you if you are:

- An individual looking for an introduction to prompt engineering
- A student aiming to graduate with practical knowledge of writing effective prompts for generative AI models, thereby increasing job readiness
- A professional seeking to enhance your professional capabilities and efficiency by leveraging the capabilities of prompt engineering in generative AI
- A manager or executive keen on using prompt engineering for strategic advantages in your organization

### Recommended Background

This course is relevant for anyone interested in exploring the field of generative AI and requires no specific prerequisites.

The course uses simple, easy-to-understand language to explain the critical concepts of generative AI without relying on technical jargon. The hands-on labs provide an opportunity

to practice the skills covered in the course and do not require a programming background or college degree.

To derive maximum learning from this course, all that's needed is active participation in and completion of the various learning engagements offered throughout the modules.

## Reading 2: Lesson Summary

At this point, you have learned the concepts of prompts and prompt engineering in generative AI. You have also explored the best practices for writing effective prompts and some common prompt engineering tools.

You learned the definition of prompts and their elements. You were introduced to prompt engineering and its relevance to generative AI models. You learned the best practices for writing effective prompts and how to refine them. You learned the functionalities and capabilities of common prompt engineering tools. You even got the opportunity to experience creating prompts and learning about naive prompting and persona patterns through hands-on lab experiences. You were privy to what experts from the field had to say about prompt engineering.

Specifically, you learned that:

- A prompt is any input or a series of instructions you provide to a generative model to produce a desired output.
- These instructions help in directing the creativity of the model and assist it in producing relevant and logical responses.
- The building blocks of a well-structured prompt include instruction, context, input data, and output indicators.
- These elements help the model comprehend our necessities and generate relevant responses.
- Prompt engineering is designing effective prompts to leverage the full capabilities of the generative AI models in producing optimal responses.
- Refining a prompt involves experimenting with various factors that could influence the output from the model.
- Prompt engineering helps optimize model efficiency, boost performance, understand model constraints, and enhance its security.
- Writing effective prompts is essential for supervising the style, tone, and content of output.
- Best practices for writing effective prompts can be implemented across four dimensions: clarity, context, precision, and role-play.
- Prompt engineering tools provide various features and functionalities to optimize prompts.
- Some of these functionalities include suggestions for prompts, contextual understanding, iterative refinement, bias mitigation, domain-specific aid, and libraries of predefined prompts.
- A few common tools and platforms for prompt engineering include IBM watsonx Prompt Lab, Spellbook, Dust, and PromptPerfect.

## **Module 2:**

### **Reading 1: Lesson Summary**

At this point, you have learned the techniques for skillfully crafting prompts that effectively steer generative AI models. You now know the various prompt engineering approaches that optimize the response of generative AI models.

You explored the techniques, including zero-shot and few-shot prompting, using which text prompts can improve the reliability of large language models (LLMs) and yield greater benefits from their responses. You learned how using different approaches such as interview patterns, Chain-of-Thought, and Tree-of-Thought to write prompts helps generative AI models produce more specific, contextual, and customized responses to the user's needs. You even had the opportunity to experience the application of each of these approaches through hands-on lab experiences. You were privy to what experts from the field had to say about the role of prompt engineering in AI.

Specifically, you learned that:

- The various techniques using which text prompts can improve the reliability and quality of the output generated from LLMs are task specification, contextual guidance, domain expertise, bias mitigation, framing, and the user feedback loop.
- The zero-shot prompting technique refers to the capability of LLMs to generate meaningful responses to prompts without needing prior training.
- The few-shot prompting technique used with LLMs relies on in-context learning, wherein demonstrations are provided in the prompt to steer the model toward better performance.
- The several benefits of using text prompts with LLMs effectively are increasing the explainability of LLMs, addressing ethical considerations, and building user trust.
- The interview pattern approach is superior to the conventional prompting approach as it allows a more dynamic and iterative conversation when interacting with generative AI models.
- The Chain-of-Thought approach strengthens the cognitive abilities of generative AI models and solicits a step-by-step thinking process.
- The Tree-of-Thought approach is an innovative technique that builds upon the Chain-of-Thought approach and involves structuring prompts hierarchically, akin to a tree, to guide the model's reasoning and output generation.

### **Reading 2: Prompt Hacks**

#### **Objective**

After completing this reading, you will be able to:

- Explain the concept of prompt hacks.
- Apply them to generate more effective outputs from generative AI models for text and image generation.

- Distinguish between text prompts and prompt engineering.

## Introduction

Prompt hacks in generative AI refer to techniques or strategies that involve manipulating the prompts or inputs provided to a generative AI model, such as a large language model (LLM) or an image generation model, to produce desired or specific outputs. These hacks include carefully crafting the prompts to influence the model's behavior and generate outputs that align with the user's intentions. They improve the performance of LLMs by:

- **Improving the quality and accuracy of LLM outputs:** By carefully crafting the prompt, you can guide the LLM toward the desired output and reduce the likelihood of errors.
- **Enabling LLMs to perform new and innovative tasks:** By combining prompts with other inputs, such as images or code, you can enable LLMs to perform tasks that they would not be able to perform otherwise.
- **Making LLMs more accessible and user-friendly:** Using prompt hacks can make it easier for people to use LLMs easily and effectively.

## Prompt hacks and text-generation

Prompt hacks can be used with LLMs to generate more effective responses from text-generating generative AI models in the following ways:

- **Using special modifiers:** LLMs have special modifiers that can control the style, tone, and other properties of the generated output. For example, you might use the modifier "write in a humorous style" to generate a funny poem or story.
- **Providing context and examples:** The more context and examples you provide to the LLM, the better it will be able to understand what you want and generate the desired output. For example, if you wish the LLM to generate a code snippet for a specific task, you might provide it with a detailed description and some examples of similar code snippets.
- **Combining prompts with other inputs:** Prompts can be combined with other inputs, such as images or code, to generate more complex and informative outputs. For example, you might prompt an LLM to generate a description of an image while providing it with the image itself.

Here's an example of using prompt hacks with text-generating models.

You need a poem about a cat. So, you use the following prompt:

**Prompt:** Write a poem about a cat.

**Output:**

A furry little creature,

With four paws and a tail,

It loves to chase mice,

And play with its ball.

Now, this is a fairly straightforward poem about a cat. However, you require a slightly innovative output. You can use prompt hacking to bring in more creativity or humor. For example, you could add the following modifier to the prompt:

**Prompt:** Write a poem about a cat in the style of a gangsta rapper.

**Output:**

Yo, I'm a cat, and I'm here to say,

I'm the coolest cat around today.

I got fur so soft and eyes so green,

And I'm the best hunter you've ever seen.

You now have a poem, about a cat, which is creative and funny!

## Prompt hacks and image generation

Did you know you can use prompt hacking with LLMs to create images?

Directing image generation models through LLMs is an emerging method that is gaining popularity in artificial intelligence. This approach entails employing an LLM to guide an image generation model, like DALL-E 2 or Imagen, to produce more precise and lifelike images.

How exactly does this work?

You provide the LLM with a text description of the image you want and then use the LLM's output to guide the image generation model.

For example, suppose you want to generate an image of a cat sitting on a couch. You could provide the LLM with a text prompt like this:

**Prompt:** "A fluffy orange cat sitting on a red couch, looking at the camera."

The LLM would then generate a response like this:

**Output:**

"A fluffy orange cat is sitting on a red couch. It is looking directly at the camera. Its eyes are green, and its fur is soft and smooth."

You can then use this response to guide the image generation model to generate an image of a cat sitting on a couch such that it matches the description provided by the LLM.

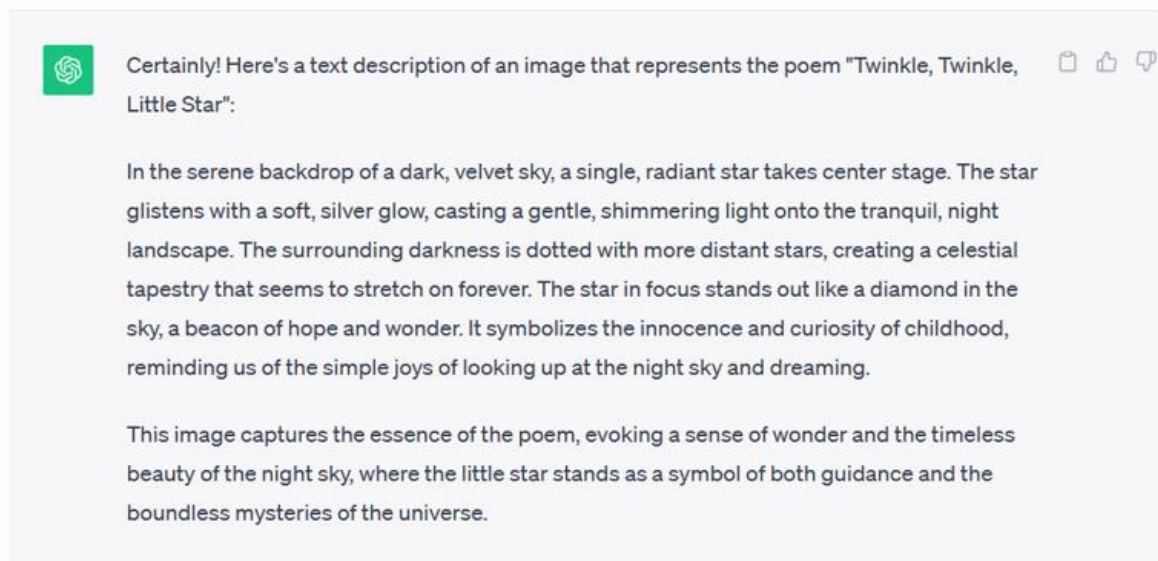
Here's another example. Imagine you need a background image for the poem "Twinkle twinkle little star."

You can use a prompt hack here and ask the LLM to suggest the prompt for generating this image.

**Prompt:** Consider the poem "Twinkle twinkle little star." Can you create a text description of an image that represents this poem?

The following image depicts the response generated by ChatGPT using this prompt.

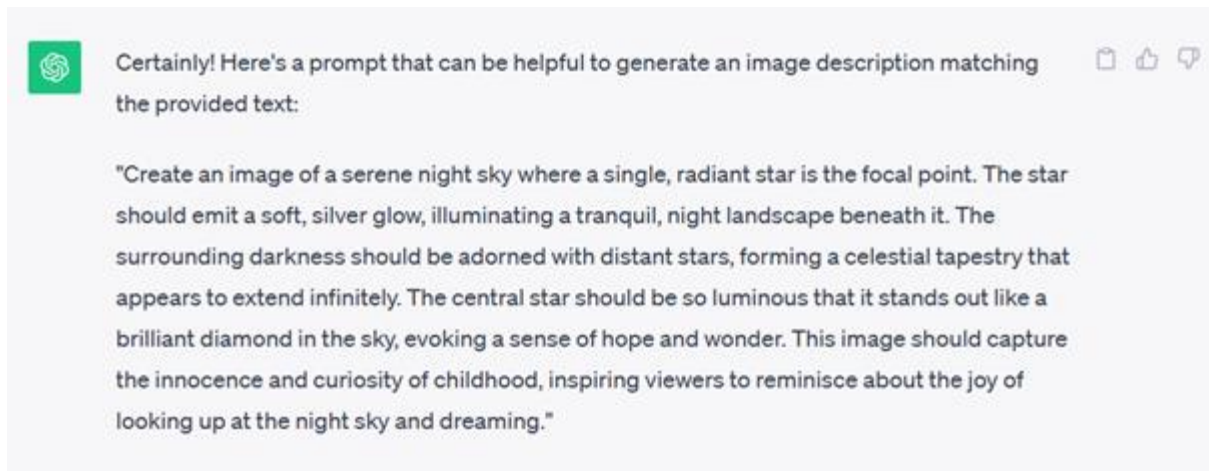
**Output:**



The prompt hack comes after this.

**Prompt:** Can you suggest a prompt that will be helpful to generate a relevant image for description: "In the serene backdrop of a dark velvet sky, a single radiant star takes center stage. The star glistens with a soft silver glow, casting a gentle, shimmering light onto the tranquil night landscape. The surrounding darkness is dotted with more distant stars, creating a celestial tapestry that seems to stretch forever. The star in focus stands out like a diamond in the sky, a beacon of hope and wonder. It symbolizes the innocence and curiosity of childhood, reminding us of the simple joys of looking up at the night sky and dreaming."

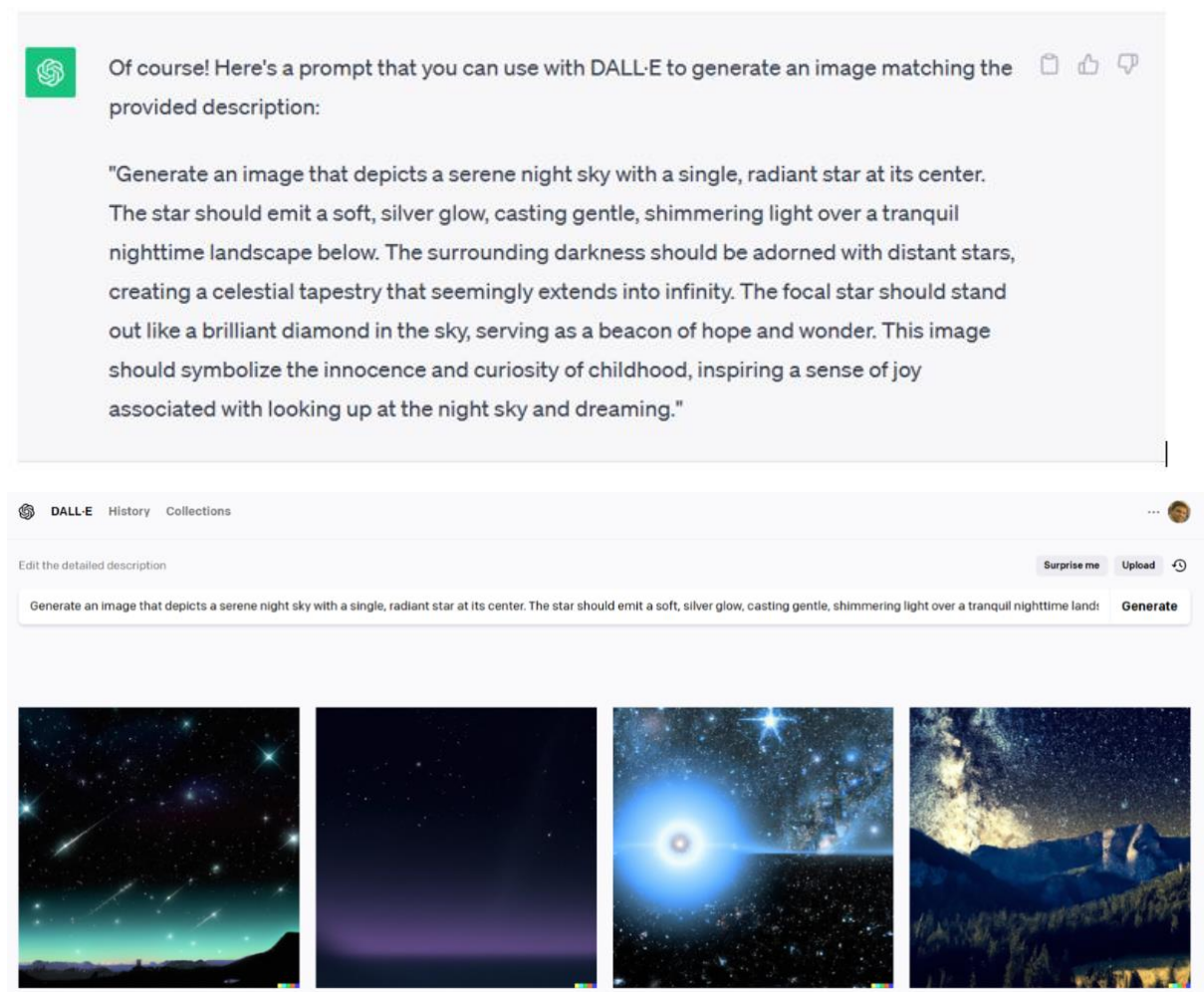
**Output:**



Now, if you need to use the same prompt for DALL-E, you can also use a prompt hack here!

**Prompt:** Can you rewrite the prompt for DALL-E?

**Output:**



Bingo!



You now have what you want. You can use these images for the poem the way you'd like.

## Prompt hacks and prompt engineering

Prompt hacking and prompt engineering are closely related fields, but they have some key differences.

Prompt hacking is the use of prompts to manipulate the output of an LLM in a way that is unexpected or unintended, whereas prompt engineering is the systematic design and development of prompts for LLMs

	Prompt hacks	Prompt engineering
Purpose	To manipulate the output of an LLM in unexpected or unintended ways	To improve the performance of an LLM on specific tasks.
Approach	Experimental and creative	Systematic and disciplined
Application	Generating humorous or creative outputs	Improving the performance of LLMs in machine translation, question answering, and other tasks

It is important to note that the distinction between prompt hacking and prompt engineering is not always clear-cut. Some techniques can be used for both purposes. For example, using special modifiers to control the style and tone of the output to generate humorous or creative outputs. It can also be used to improve the performance of an LLM on a specific task, such as generating text in a specific style.

## Tips for powerful prompt hacking

Here are some additional tips for prompt hacking:

- Be creative, and don't be afraid to try new things
- Be specific and clear in your instructions.
- Use the LLM's documentation to learn more about its capabilities and limitations.
- Experiment with different prompts and see what works best for you.

With some practice, you can use prompt hacking to generate high-quality and creative outputs from LLMs.

To conclude, prompt hacking is a powerful technique that can be used to get the most out of LLMs. However, it is a relatively new field, and there is no one-size-fits-all approach. The best way to learn how to hack prompts is to experiment and discover what works for you.

## Summary

You learned the concept of prompt hacks in generative AI. You also learned how they can be used with LLMs for better text and image generation. Finally, you learned the difference between prompt hacking and prompt engineering.

Module 3

## Reading 3: Generative AI: Prompt Engineering

Welcome! This alphabetized glossary contains many of the terms in this course. This comprehensive glossary also includes additional terms not used in course videos. These terms are essential for you to recognize for better comprehension of the concepts covered in the course.

Term	Definition
API integration	Application programming interface integration refers to the process of connecting different software systems or applications through their APIs to enable them to work together and share data or functionality.
Bias mitigation	A technique in which text prompts provide explicit instructions to generate neutral responses.
Chain-of-Thought	An approach to prompt engineering that involves breaking down a complex task into smaller and easier ones through a sequence of more straightforward prompts.
ChatGPT	A language model designed to provide detailed responses to natural language input.
Claude	A powerful and flexible AI chatbot to help you with your tasks.
Contextual guidance	A technique using which text prompts provide specific instructions to the LLMs to generate relevant output.
DALL-E	Text-to-image model that generates digital images from natural language descriptions.

Term	Definition
Domain expertise	A technique wherein text prompts can use domain-specific terminology to generate content in specialized fields like medicine, law, or engineering, where accuracy and precision are crucial.
Dust	A prompt engineering tool that provides a web user interface for writing prompts and chaining them together.
Explainability	Refers to the degree to which a user can understand and interpret the model's decision-making process and the reasons behind its generated outputs.
Few-shot prompting	A method that enables context learning, wherein demonstrations are provided in the prompt to steer the model to better performance.
Framing	A technique by which text prompts guide LLMs to generate responses within the required boundaries.
Generative AI	A type of artificial intelligence that can create new content, such as text, images, audio, and video.
Generative AI models	Models that can understand the context of input content to generate new content. In general, they are used for automated content creation and interactive communication.
GPT	Generative pre-trained transformers or GPT are a family of neural networks that uses transformer architecture to create human-like text or content as output.
IBM watsonx.ai	A platform of integrated tools to train, tune, deploy, and manage foundation models easily.
Integrated Development Environment (IDE)	A software tool for crafting and executing prompts that engage with language models.
Input data	Any piece of information provided as part of the prompt.
Interview pattern approach	A prompt engineering strategy that involves designing prompts by simulating a conversation or interacting with the model in the style of an interview.

Term	Definition
LangChain	A Python library that provides functionalities for building and chaining prompts.
Large language models (LLMs)	A type of deep learning model trained on massive amounts of text data to learn the patterns and structures of language. They can perform language-related tasks, including text generation, translation, summarization, sentiment analysis, and more.
Midjourney	A text-to-image model that generates images from natural language requests.
Naive prompting	Asking queries from the model in the simplest possible manner.
Natural language processing (NLP)	A branch of artificial intelligence that enables computers to understand, manipulate, and generate human language (natural language).
OpenAI Playground	A web-based tool that helps to experiment and test prompts with various models of OpenAI, such as GPT.
Output indicator	Benchmarks for assessing the attributes of the output generated by the model.
Prompt	Instructions or questions given to a generative AI model to generate new content.
Prompt engineering	The process of designing effective prompts to generate better and desired responses.
PromptBase	A marketplace for selling and buying prompts.
Prompt lab	A tool that enables users to experiment with prompts based on different foundation models and build prompts based on their needs.
PromptPerfect	A tool used to optimize prompts for different LLMs or text-to-image models.
Role-play/Persona pattern	Specific format or structure for constructing prompts that involve the perspective of a character or persona.
Scale AI	A technology company that specializes in data labeling and data annotation services.

Term	Definition
Stable Diffusion	A text-to-image model that generates detailed images based on text descriptions.
StableLM	An open-source language model based on a dataset that contains trillions of tokens of content.
Tree-of-Thought	An approach to prompt engineering that involves hierarchically structuring a prompt or query, akin to a tree structure, to specify the desired line of thinking or reasoning for the model.
User feedback loop	A technique wherein users provide feedback to text prompts and iteratively refine them based on the response generated by the LLM.
Zero-shot prompting	A method using which generative AI models generate meaningful responses to prompts without needing prior training on those specific prompts.

## Reading 4: Data Labeling & Data Annotation Services

Both **data labeling** and **data annotation** refer to the process of adding meaningful metadata to raw data (images, videos, text, audio, etc.) to make it usable for machine learning models.

### 1. Data Labeling

It involves categorizing and tagging data with predefined labels so that an AI model can learn from it. Data labeling is essential in supervised learning, where models require labeled datasets to make accurate predictions.

- **Example:** Labeling images of cats and dogs as "cat" or "dog" for an image classification model.

### 2. Data Annotation

A more detailed process of enriching raw data with additional information (bounding boxes, segmentation masks, keypoints, sentiment labels, etc.) to enhance AI training.

- **Example:** In self-driving car datasets, annotating pedestrians, traffic signs, and lane markings with bounding boxes or polygons.

## Types of Data Annotation Services

Depending on the type of AI application, data annotation services can be:

1. **Image Annotation**
  - **Bounding Boxes** – Drawing rectangular boxes around objects.
  - **Semantic Segmentation** – Labeling each pixel of an image.
  - **Instance Segmentation** – Assigning unique labels to individual objects in an image.
  - **Keypoint Annotation** – Marking important points (e.g., facial landmarks).
2. **Text Annotation**
  - **Named Entity Recognition (NER)** – Identifying entities like names, locations, or brands.
  - **Sentiment Analysis** – Classifying text based on emotions (positive, negative, neutral).
  - **Part-of-Speech (POS) Tagging** – Labeling words as nouns, verbs, adjectives, etc.
3. **Audio Annotation**
  - **Speech-to-Text Transcription** – Converting spoken words into text.
  - **Speaker Diarization** – Identifying different speakers in an audio clip.
  - **Emotion Recognition** – Analyzing voice tone and sentiment.
4. **Video Annotation**
  - **Object Tracking** – Tracking moving objects across frames.
  - **Action Recognition** – Labeling specific actions like running, jumping, or waving.

Who Needs Data Annotation Services?

- **AI & ML Startups** – For training models in computer vision, NLP, and speech recognition.
- **Autonomous Vehicle Companies** – To improve self-driving algorithms.
- **Healthcare & Medical AI** – For annotating medical scans (X-rays, MRIs).
- **E-commerce & Retail** – To enhance product recommendations and search engines.

## Reading 5: About the Optional Lesson using IBM watsonx.ai

In this optional lesson you will explore the Prompt Lab in IBM watsonx.ai. This is optional because it requires you to access IBM watsonx platform, which may not be available to everyone. Normally a credit card is required to access watsonx.ai. However, for learners in this course we are providing you with a feature code that will enable you to sign up without a credit card. To complete the lab you will need to perform the following:

1. Obtain a feature code for an IBM Cloud account.
2. Create the IBM cloud account using the feature code.
3. Provision your instance for watsonx.ai and explore the Prompt Lab.

## Reading 6: Activate Trial Account

To facilitate hands-on skills development and enable you to complete some of the labs in this course, we are providing you with a special feature code for IBM Cloud.

This feature code will enable you to either create an IBM Cloud account without a credit card.

Please check the box and click on the "**Launch App**" button below to get a unique Feature Code to activate your IBM Cloud trial account.

We have also provided you with the stepwise instructions on how to create an IBM Cloud account for your reference.

**NOTE:** If you have already applied your IBM Cloud feature code in another course/lab to create an IBM Cloud account, please skip this item, as the code can only be used once.

**CAUTION:** While signing up for an IBM Cloud account, or at a later stage, you may be prompted to enter a credit card. Please note that access to paid services or a credit card is **NOT** required to complete any mandatory labs/projects / or assignments in this course. Instead, those enrolled in this course have access to a **Feature Code**, which enables trial access to IBM Cloud for a limited duration. Therefore, we strongly recommend that you **DO NOT enter a credit card** to complete the required course components.

If you still choose to enter your Credit Card to use IBM Cloud services, you will be liable to pay any resulting charges billed to your credit card. The course team does not take any responsibility for such charges.

In case you encounter issues while activating your account with a Feature Code, or your trial has ended, and you have not yet completed the course, please contact the course support team via the course discussions forums for assistance instead of adding a credit card to your IBM Cloud account.

This course uses a third-party app, Obtain IBM Cloud Feature Code and Activate Trial Account, to enhance your learning experience. The app will reference basic information like your name, email, and Coursera ID.

Trial Account Pin: **b4a668149883ec601d50793886f744ba**

## Exploring IBM watsonx Prompt Lab

When it comes to prompting foundation models, there isn't just one right answer. There are usually multiple ways to prompt a foundation model for a successful result.

The **IBM watsonx Prompt Lab** is a prompting tool designed to help you craft prompts that achieve the desired or expected result from a given model and set of parameters. In this activity, you will explore the key features of the watsonx Prompt Lab, and also learn how to prompt a foundation model.

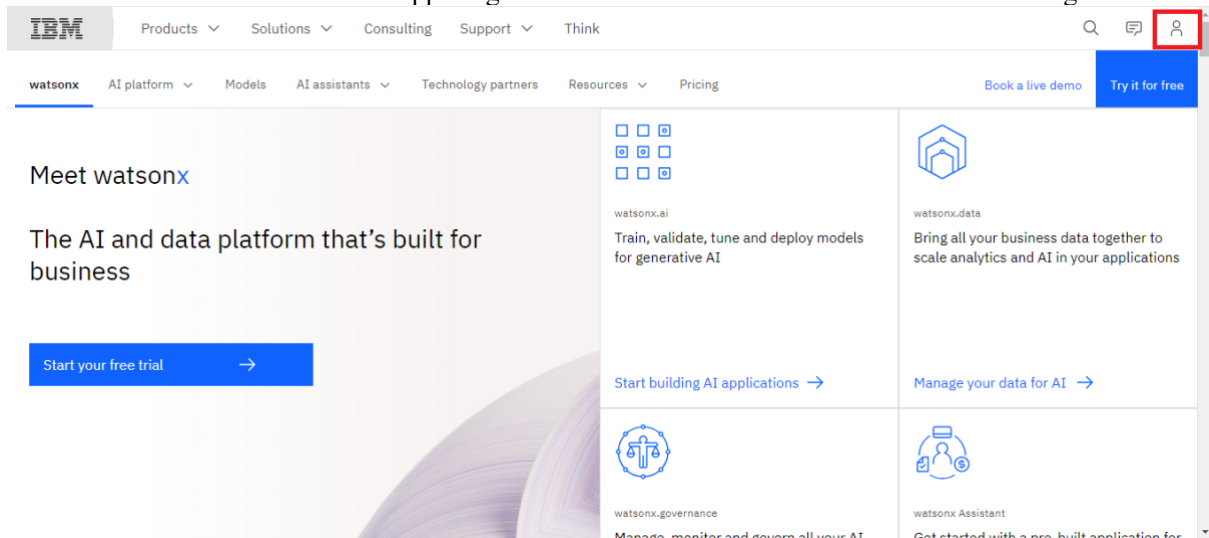
## Accessing IBM watsonx

To access the IBM watsonx platform, you need to have an IBM cloud account. You can use the credentials for the IBM cloud account to access the watsonx platform.

If you do not have an IBM cloud, you will be asked to register and create your account when you try to access the IBM watsonx platform.

Here is the link to the platform: [IBM watsonx](https://www.ibm.com/watsonx).

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.






# Log in to IBM

IBMid

Continue



☐ Remember me 

Don't have an account?

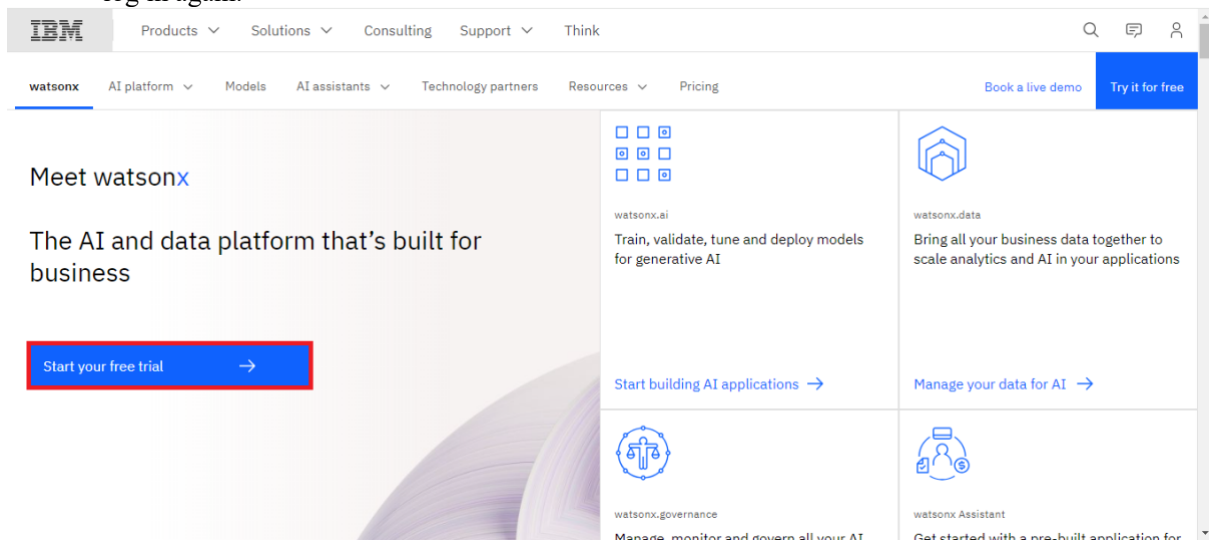
Create an IBMid



Forgot IBMid? [Contact the IBMid help desk](#)

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

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



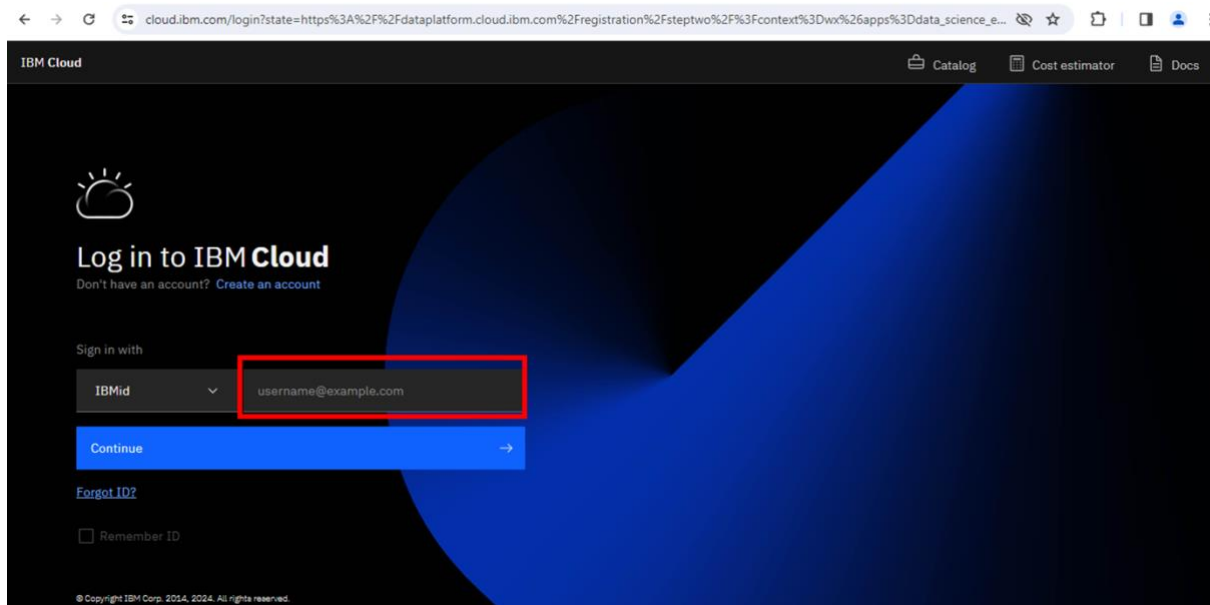
The screenshot shows the IBM watsonx home page. The top navigation bar includes the IBM logo, links for Products, Solutions, Consulting, Support, and Think, a search icon, and user profile icons. Below this is a secondary navigation bar with links for watsonx, AI platform, Models, AI assistants, Technology partners, Resources, and Pricing. On the right side of this bar are links for 'Book a live demo' and 'Try it for free'. The main content area features a large hero section on the left with the text 'Meet watsonx' and 'The AI and data platform that's built for business', followed by a red-bordered button that says 'Start your free trial' with a right arrow. To the right of the hero section is a grid of four cards: 1. 'watsonx.ai' with a grid icon and text 'Train, validate, tune and deploy models for generative AI', with a link 'Start building AI applications'. 2. 'watsonx.data' with a hexagon icon and text 'Bring all your business data together to scale analytics and AI in your applications', with a link 'Manage your data for AI'. 3. 'watsonx.governance' with a scales icon and text 'Manage, monitor and govern all your AI'. 4. 'watsonx Assistant' with a person icon and text 'Get started with a pre-built application for...'. The bottom of the page shows a partial view of a large, abstract, purple-toned graphic.

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. Choose the same region you had used to create your account.

The first screenshot shows the IBM WatsonX AI registration page. The URL is `dataplatfom.cloud.ibm.com/registration/stepone?context=wx&preselect_region=true`. The page has a dark theme and says "Try IBM watsonx.ai for free". It prompts the user to "Build, train, validate, tune and deploy AI models." Under "Select a region", "Dallas (us-south)" is selected. At the bottom, there are two buttons: "Create account or log in" and "Log in". The "Log in" button is highlighted with a red box.

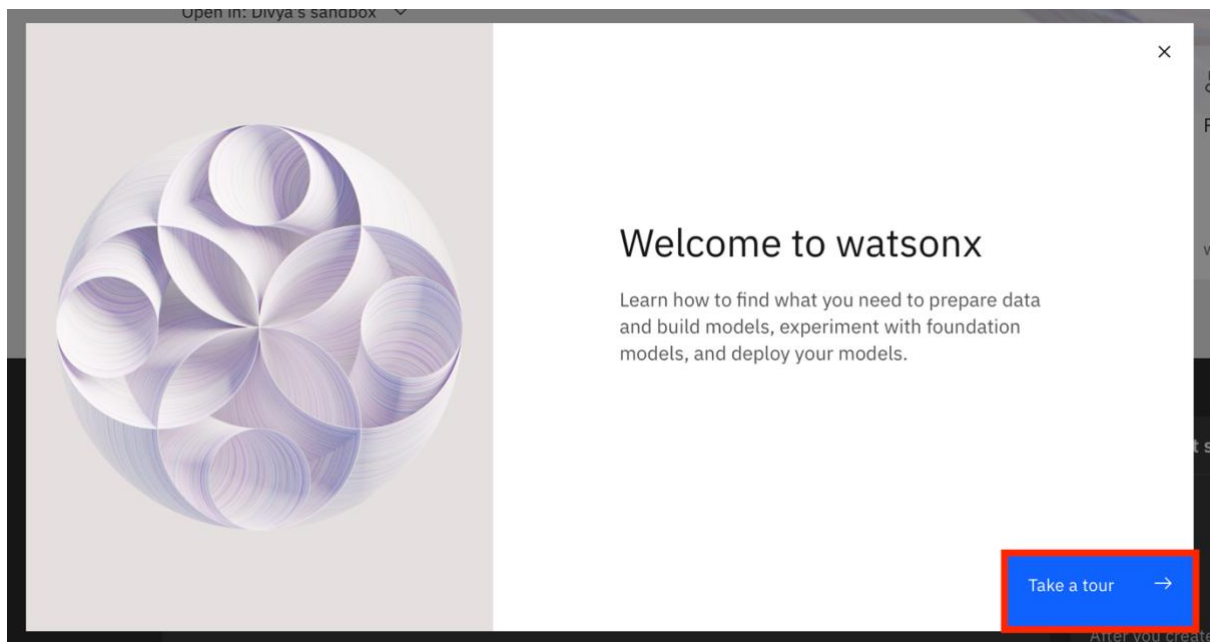
The second screenshot shows the IBM Cloud login page. The URL is `cloud.ibm.com/registration?target=https%3A%2F%2Fdataplatfom.cloud.ibm.com%2Fregistration%2Fsteptwo%2F%3Fcontext%3Dwx%26apps%3Ddata_sci...`. It says "Create an IBM Cloud account" and "Already have an IBM Cloud account? Log in". The "Log in" link is highlighted with a red box. Below this is a form with fields for "Email" and "Password", and a "Next" button. To the right, there is a large banner for "watsonx.ai" with the text "Get started for free with watsonx.ai™" and "Build and deploy generative AI and machine learning solutions."

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



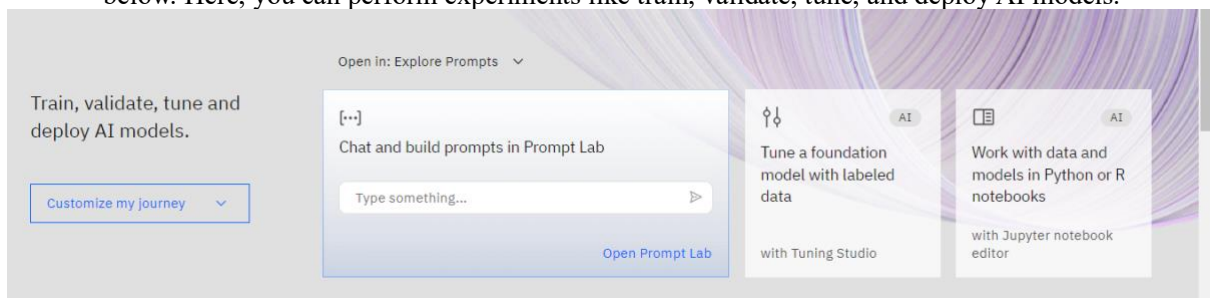
7. Provide the required information and then click on **Continue**.

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



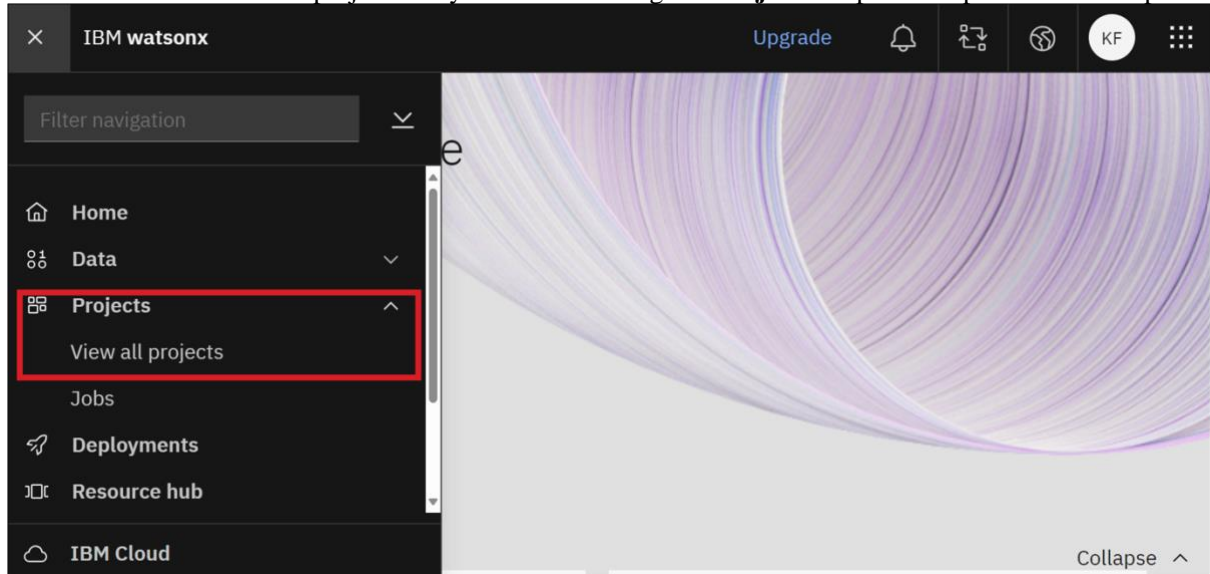
*Note: You may opt for the watsonx walkthrough by clicking **Take a tour** or skip or exit it by clicking **X** at the upper right corner of the screen*

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.

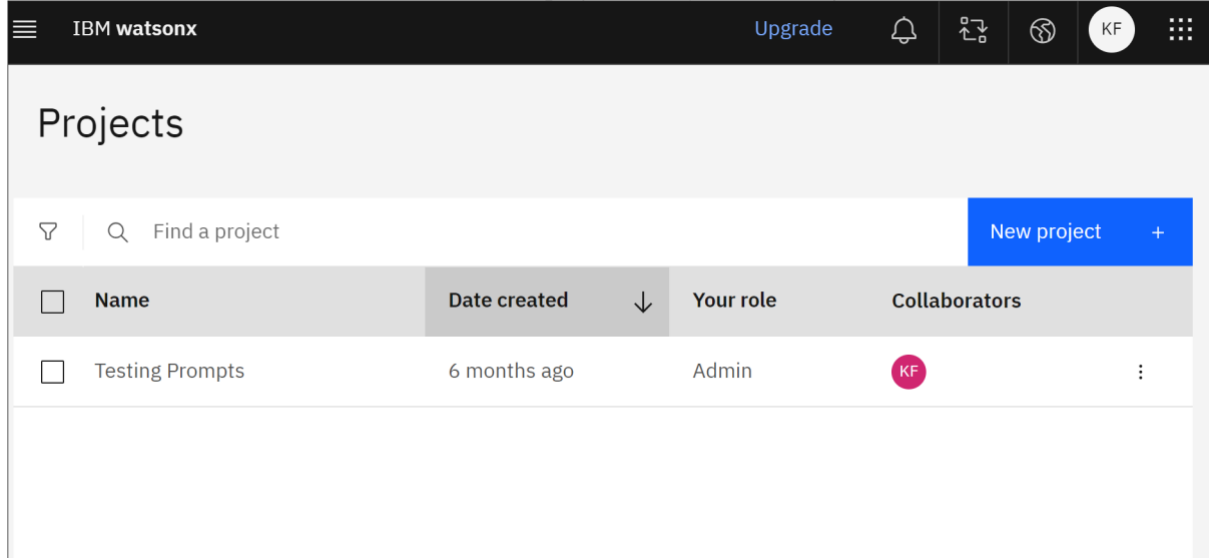


*Note: If this is NOT your first login session in the watsonx.ai, you may find the **Prompt Lab** deactivated. In this case, you need to create a project and then select the **Open Prompt Lab** option. You can create a new project by following the given steps.*

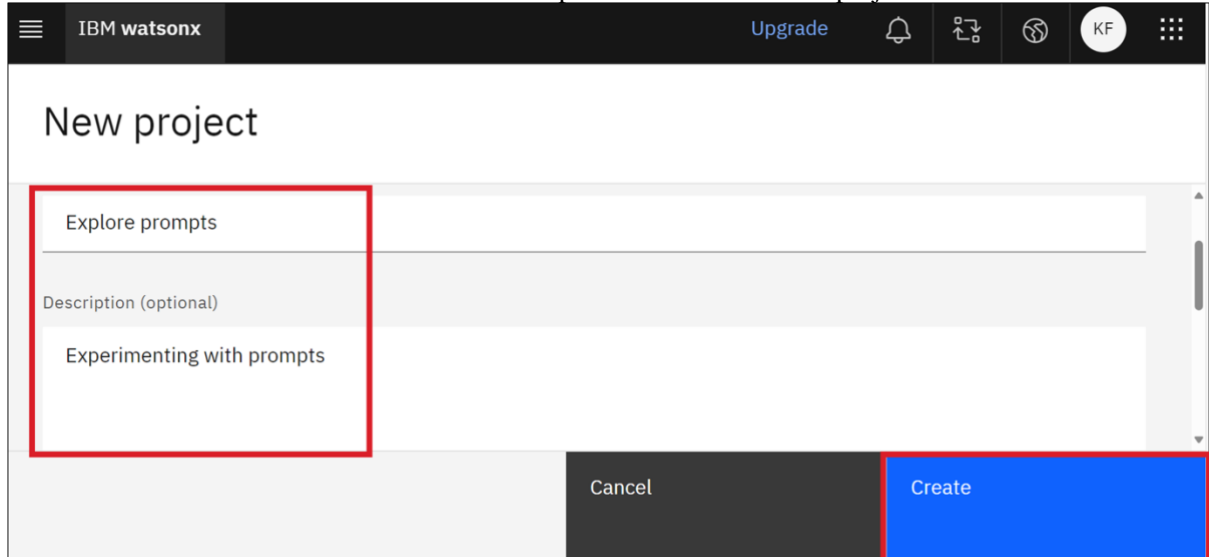
- a. You can view the list of projects for your account using the **Projects** drop-down option in the left panel.



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

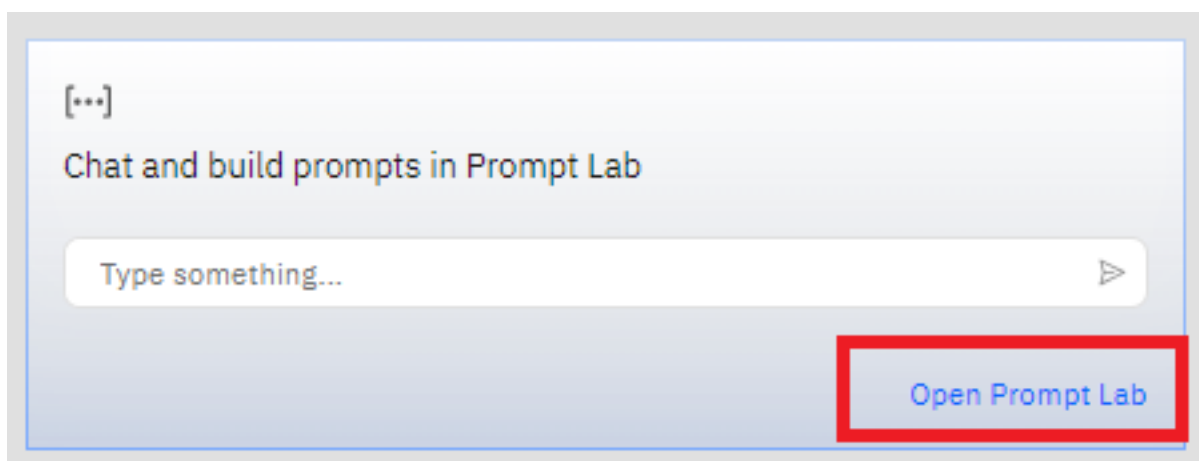


c. Enter the name and description of the project and click **Create**.



d. Go to **Home** using drop-down option in the left panel.

e. You will reach your IBM watsonx.ai dashboard page. Click the first card with Prompt Lab, as shown in the screenshot below.



f. Click on **Associate Service**

## No Watson Machine Learning service detected

To access the Prompt Lab, you must associate a Watson Machine Learning service instance to this project.

Associate service



g. Select **Watson Machine Learning-vy** and click **Associate**. If Watson Machine Learning is not seen please select **New Service** and add **Watson Machine Learning**

Associate service

Choose an existing or add a new service to associate with your project.

1 X Default 2 X Locations

Find services [New service +](#)

Name	Type	Plan	Location	Status	Group
<input checked="" type="checkbox"/> Watson Machine Learning-vy ⓘ	Watson Machine Learning	Lite	Dallas	Not associated	Default

Cancel Associate

h. Go to **Home** using drop-down option in the left panel.

Please be aware that due to specific geographic and business constraints, some may not be eligible to register. If you encounter this limitation, you can still explore the vast landscape of Large Language Models through alternative services such as ChatGPT.

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

[...]

Chat and build prompts in Prompt Lab

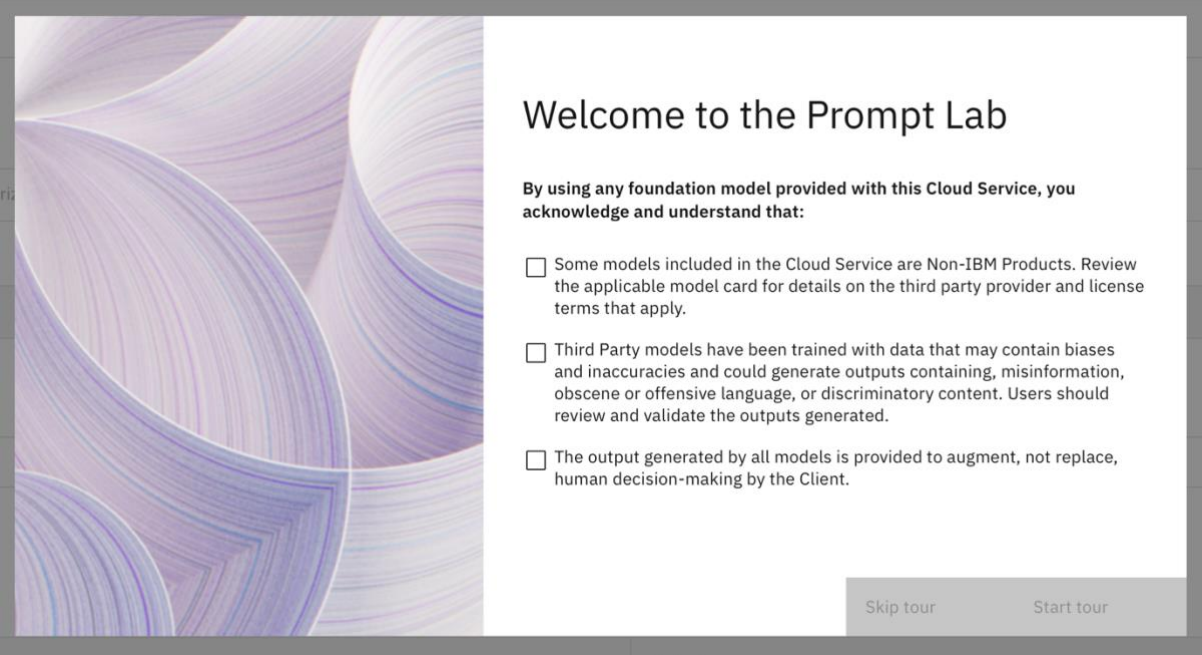
Type something...

Open Prompt Lab



*Note: If you have created a new project, click **Open Prompt Lab** in the Start working section*

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



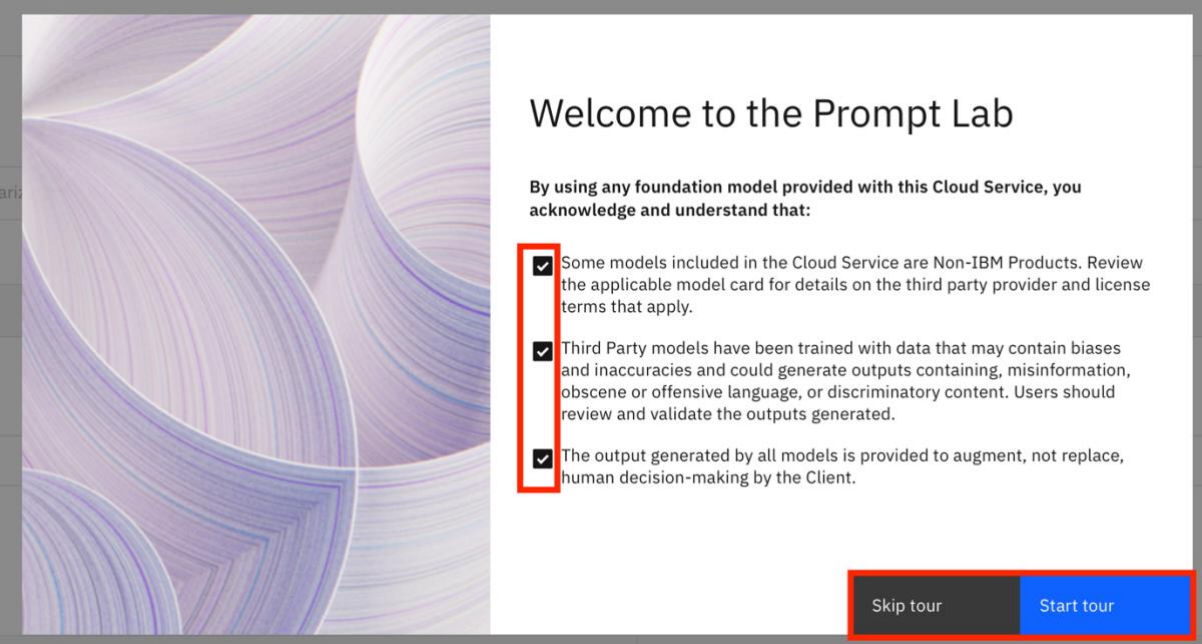
## Welcome to the Prompt Lab

By using any foundation model provided with this Cloud Service, you acknowledge and understand that:

- ☐ Some models included in the Cloud Service are Non-IBM Products. Review the applicable model card for details on the third party provider and license terms that apply.
- ☐ Third Party models have been trained with data that may contain biases and inaccuracies and could generate outputs containing, misinformation, obscene or offensive language, or discriminatory content. Users should review and validate the outputs generated.
- ☐ The output generated by all models is provided to augment, not replace, human decision-making by the Client.

Skip tour Start tour

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



## Welcome to the Prompt Lab

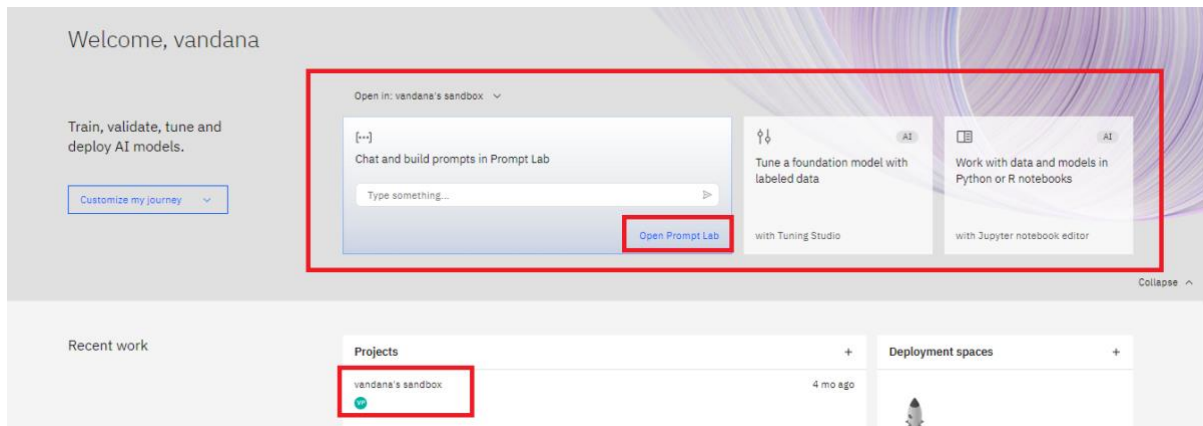
By using any foundation model provided with this Cloud Service, you acknowledge and understand that:

- ☒ Some models included in the Cloud Service are Non-IBM Products. Review the applicable model card for details on the third party provider and license terms that apply.
- ☒ Third Party models have been trained with data that may contain biases and inaccuracies and could generate outputs containing, misinformation, obscene or offensive language, or discriminatory content. Users should review and validate the outputs generated.
- ☒ The output generated by all models is provided to augment, not replace, human decision-making by the Client.

Skip tour Start tour

## Getting Started with watsonx

The Home page should look similar to this screen.



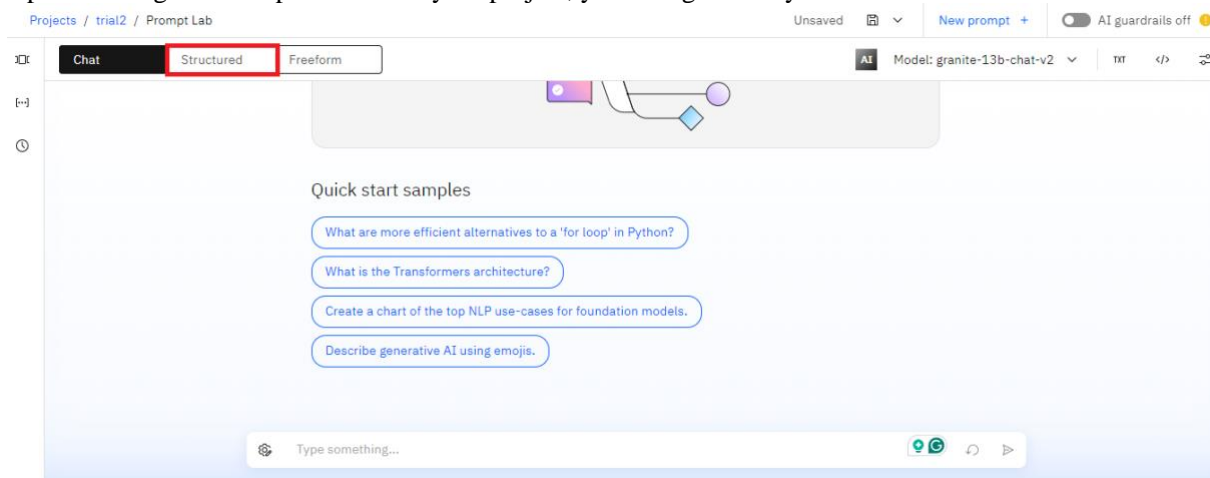
If you want to work with the Prompt Lab, you will need to work in the context of a project. If you look carefully, you will notice a highlighted square at the bottom of this screen. That's your personal sandbox. In this instance, it's labeled "Vandana's sandbox", as Vandana has signed in to her watsonx account.

Within each project, a suite of tools awaits, enabling you to interact with the AI, construct machine learning models, utilize Jupyter notebooks, and more. Of course, you're welcome to create more projects to organize and structure your work.

If you wish to do that, choose a project from the watsonx.ai home page, and click **Open Prompt Lab**. This will open the Prompt Lab.

## Part A - Exploring the Prompt Lab

Upon entering the Prompt Lab within your project, you'll be greeted by this interface. Select **Structured** tab.



Familiarize yourself with the tool's features. On the left, you'll find icons representing Sample Prompts, Saved Prompts, and History. By default, these are collapsed.



Additionally, observe the options to create a New Prompt, save your current work, and turn on or off the AI Guardrails. These guardrails are essential safety measures designed to prevent the AI from producing harmful or offensive content. Feel free to explore these features.

Once you've had a good look around, let's delve deeper into the tool's primary components.

At the top, you'll spot three tabs: **Chat**, **Structured** and **Freeform**. As the name suggests, the Structured tab offers a guided approach to input, reminiscent of a more advanced version of the AI Classroom you've previously used in this course. On the other hand, the Freeform tab presents a simple textbox, allowing you to directly query the AI.

Take time to familiarize yourself with the Structured mode, as it offers a more guided experience.

There are three sections for you in here:

1. **Instruction Section:** This is akin to our earlier "Prompt Instructions."
2. **Example Section:** Here, you can provide specific input-output pairs to guide the AI, similar to our exercises in the Chain-of-Thought lab.
3. **Try Section:** This is where you'll input your queries. The AI's responses will be displayed in the Output field within this section.

Finally, note the **Generate** button located at the bottom right. Initially, it appears grayed out, indicating that no input has been provided. Once you enter your query, the button will become active, allowing you to submit your input to the AI for processing.

## Exercise 1: Getting Started with Prompt Lab

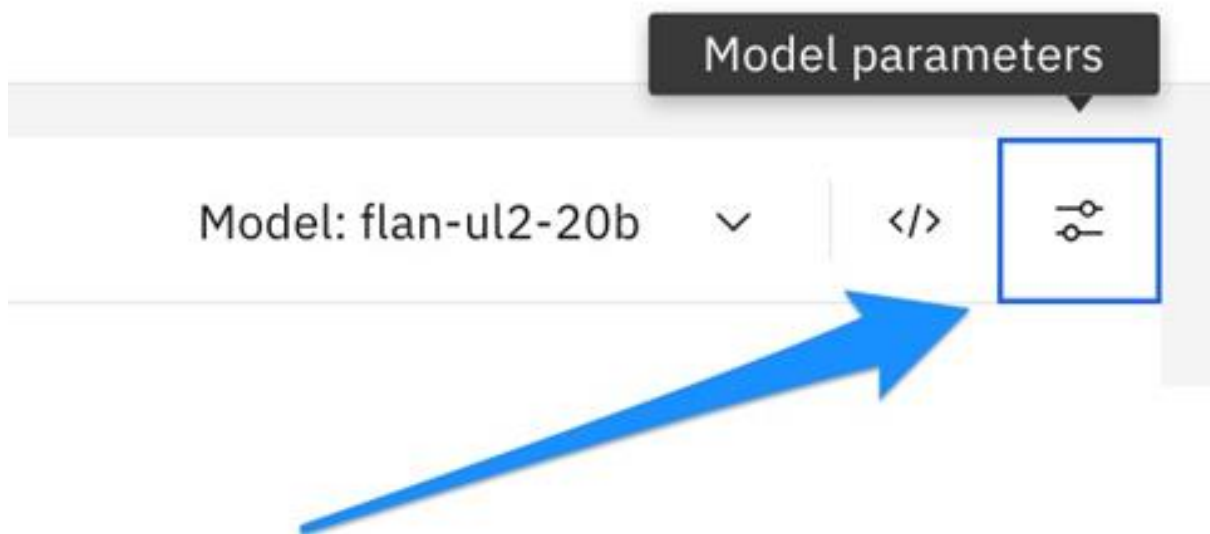
The following exercises should familiarize you with the Prompt Lab interface. For future convenience, consider saving prompts which you find particularly useful. You can use them as templates, simply tweaking the input field to suit your current query. This method is both efficient and time-saving. Select model **flan-ul2-20b**

1. **Basic Query:** Without any prompt instructions or examples, ask the AI:

1. Give me a list of common fish species in North America.

You might find the output lacking. For instance, I received `List of fish species in North America`. Where's the list? This truncation occurs in part because the default chat setup limits the response to a set word count. Let's rectify this.

2. **Adjusting Model Parameters:** Click the Model parameters settings icon in the top right, as depicted below.



Adjust the Max token parameter within a range of 1 to 1024. For this exercise, set it to 200, capping our output at 200 words. You can also set a minimum word count and tweak more advanced parameters. Remember, the cost of using foundation models in IBM WatsonX is determined by usage, which relates to the number of tokens generated. Therefore, setting a conservative Max tokens value is cost-effective.

## Model parameters

### Decoding

Greedy ☒ Sampling ⓘ

Repetition penalty

1

●

2

1

### Stopping criteria ⓘ

Stop sequences

+

Min tokens

0

Max tokens

200

After setting the 200-token limit, close the

Model parameters and click Generate again. If the output remains unsatisfactory, don't worry. We have more tricks up our sleeve!

- Switching Models:** From the dropdown adjacent to Model parameters, switch from the `flan-ul2` model to IBM's `mixtral-8x7b-instruct-v01-q` or `granite-13b-chat-v2` model. Click Generate. You should now receive a comprehensive list of common fish species in North America. It doesn't apply here, but it's worth noting that if you're pleased with the output, you can also use the top right corner of the Output field to transfer the output to the Example section. This ensures future responses are modeled after this successful example.
- Guardrail Safeguards:** You might observe that one entry was omitted due to Guardrail safeguards. Temporarily disable the guardrail at the page's top and generate a new output. Can you identify which panfish species triggered the censor?
- Improving English Clarity:** Modify the instructions to state:

```

1. 1
1. Acting as an English professor, you'll take my input, enhance its clarity, and make it sound more native.

```

Then, input the following fractured English sentence:

1. Yesterday, I go to big market for buy many fruits, but they no have the apple I like, so me feel sad and come back home without buy anything.

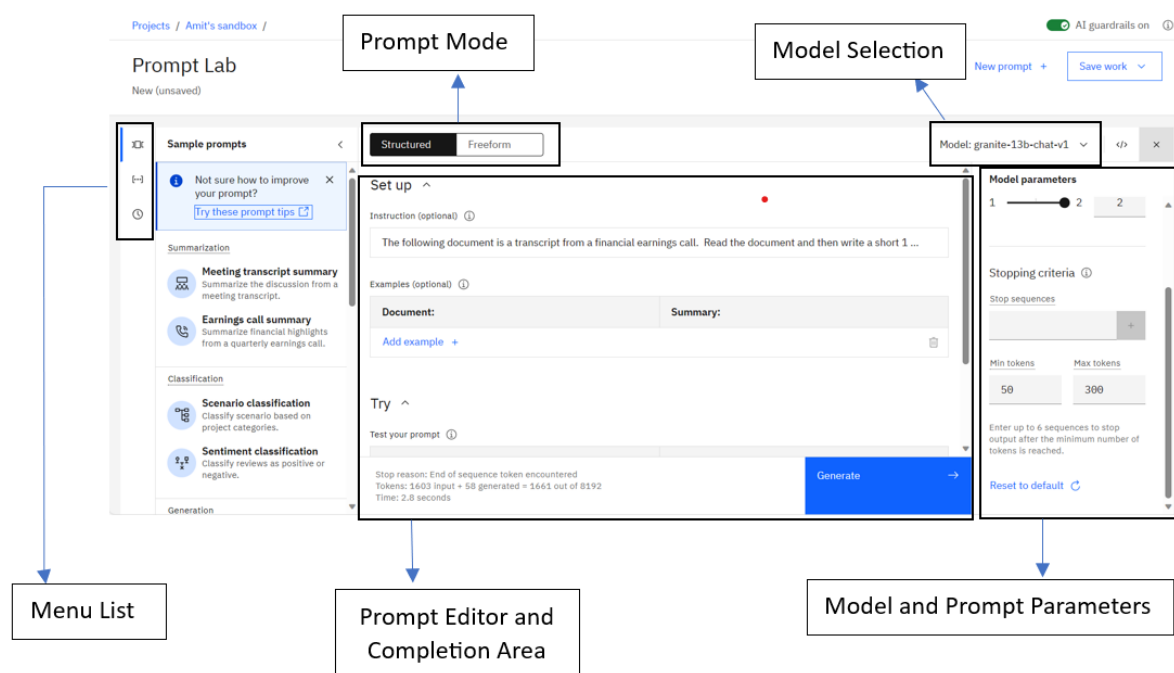
Copied!

Experiment with various models. The results should be impressive, making this tool invaluable for non-native English speakers.

6. **Translation:** Use the prompt instructions to guide the model to translate your input into Italian or a second language you might speak. Then, input an English sentence. If you do speak a second language, try it with a few different models to see which one performs best. In my case, LLama by Meta did a particularly impressive job with Italian. watsonx is also smart enough to recommend when a specific prompt can benefit from an example, guiding you toward better results.

## Part B - Prompting a foundation model

We can now explore in further detail, a few of the functions available on the watsonx Prompt Lab page with a focus on querying foundation models.



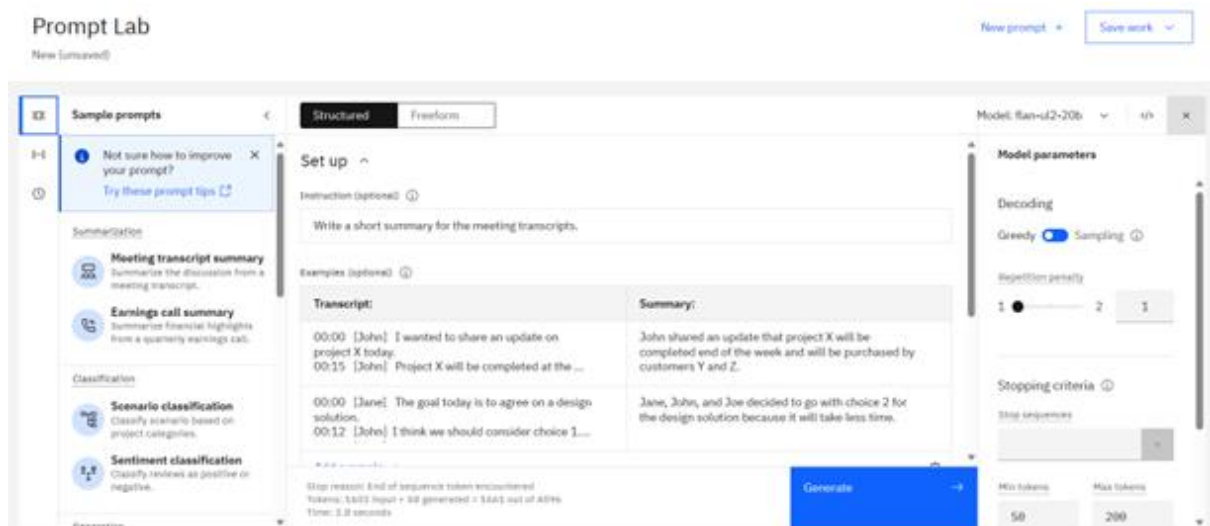
## Menu List

The menu lists Sample prompts, which you can use. All you need to do is select the task-specific scenario, and the corresponding prompt will appear in Prompt Editor. This is a good way to learn how to prompt a foundation model.

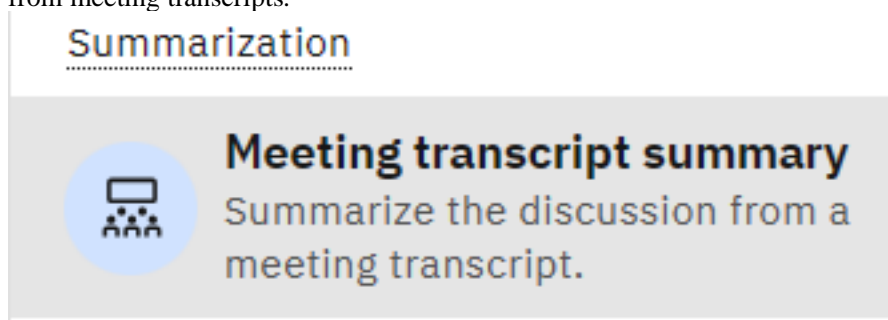
Instructions to use:



Click on sample prompt icon. Sample prompt may be expanded on the left side of Prompt Editor.



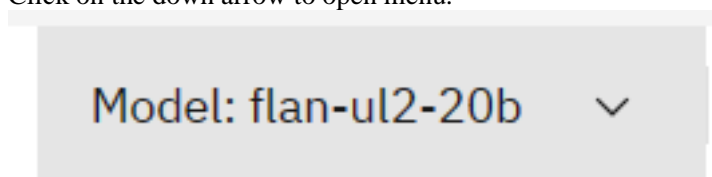
If you click on **Meeting transcript summary**, you will see the sample prompts listed for summarizing data from meeting transcripts.



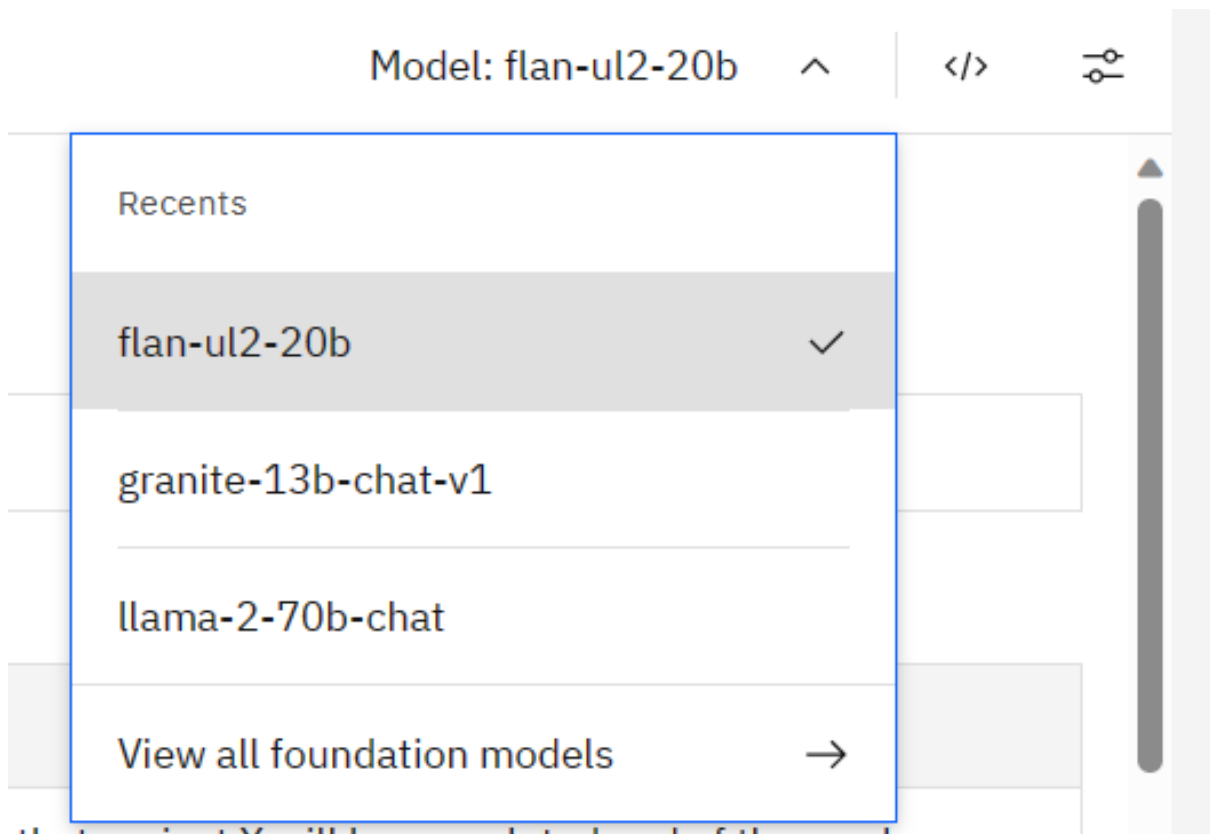
Under **Classification**, you will find sample prompts to help you sort data based on your preferences.

## Model Selection

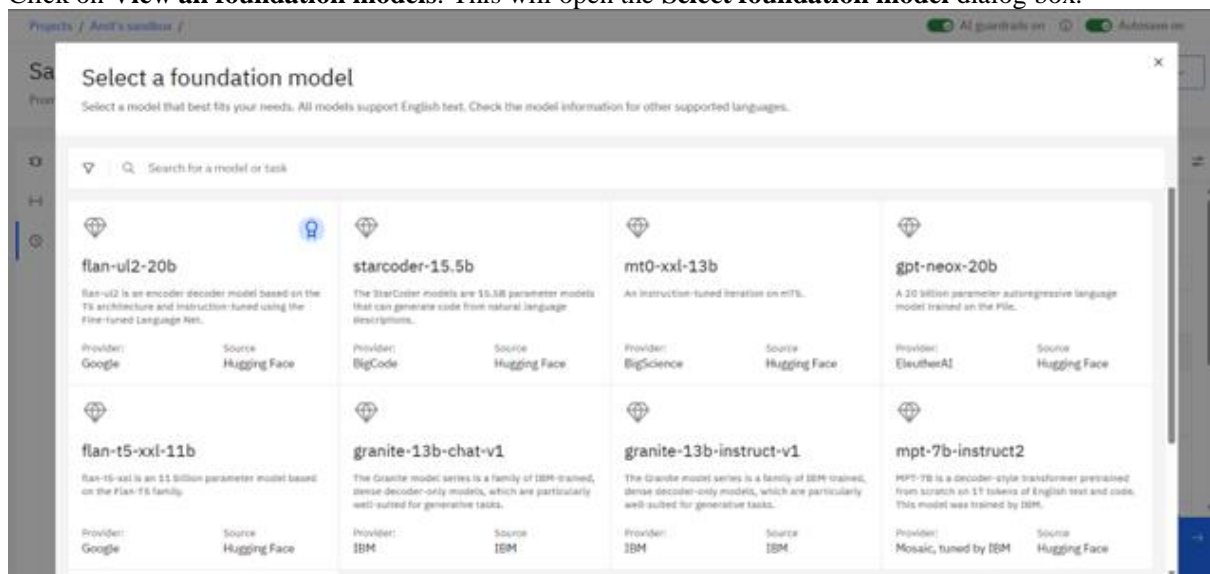
Another interesting feature is the Model Selection button. Click on the down arrow to open menu.



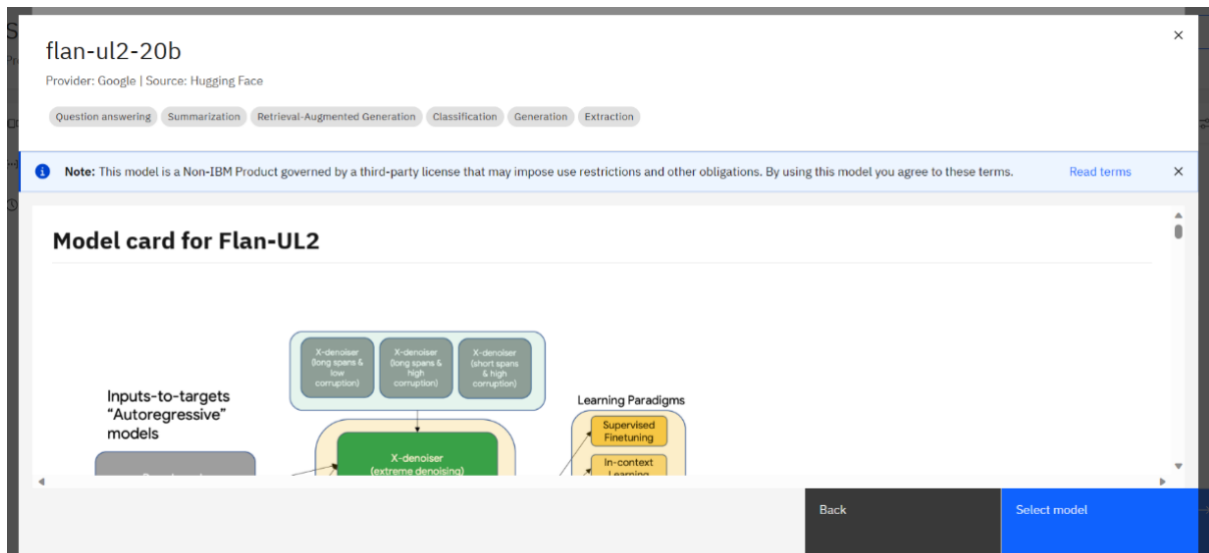
It will list some foundation models and allow you to explore more.



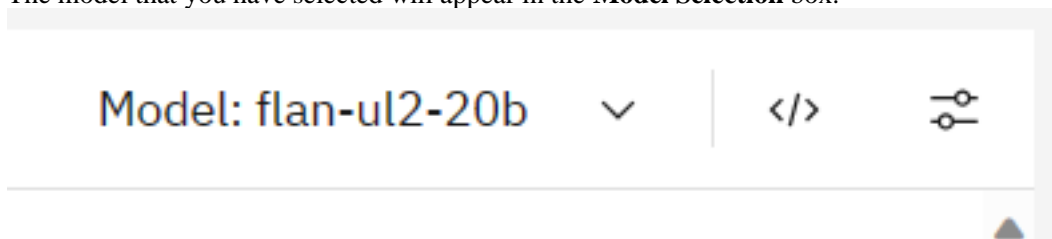
Click on **View all foundation models**. This will open the **Select foundation model** dialog box.



Click on any model. This will open a model card or **Model Datasheet** that contains detailed information about the model.



To select a model, click on the **Select Model** button at the bottom right corner of the dialog box. The model that you have selected will appear in the **Model Selection** box.



## Model and Prompt Parameters

The Model and Prompt Parameters function allows you to specify parameters to control how the model generates output in response to your prompt. Here, you will find a list of parameters you can control in the Prompt Lab.

### Decoding

Decoding is the process a model uses to choose the tokens in the generated output. Greedy decoding selects the token with the highest probability at each step of the decoding process. Sampling decoding is more variable and random, which is desirable in creative use cases. However, with greater variability comes the risk of nonsensical output.

### Repetition

If you notice the result generated for your chosen prompt, model, and parameters consistently contains repetitive text, you can try adding a repetition penalty. The higher the penalty, the less likely the result will include repeated text.

### Stopping

You can affect the length of the output generated by the model in two ways: Specifying stop sequences and setting Min tokens and Max tokens. Min tokens must be less than or equal to Max tokens.

### Penalty

### Criteria

## Exercise 2: How to prompt a foundation model

### Scenario

You work with a cloud software company, and a customer-support chatbot has received a message from a customer. You are tasked with classifying the customer's message as a question or a problem description so the chat can be routed to the correct support team. Let's see how you can adjust the model parameters to achieve this task.

### Step

Click on the **New Prompt** button

- Step** 2  
Select a model: either mt0-xxl-13b, flan-t5-xxl-11b, or flan-ul2-20b  
The model must only return one of the class names, not be creative and make up new classes.
- Step** 3  
Select the decoding as Greedy.
- Step** 4  
The Repetition penalty is set to 1 by default.
- Step** 5  
Select the Stopping criteria. Specify two stop sequences: "Question" and "Problem". After the model generates either of those words, it should stop.
- Step** 6  
Set the Max tokens parameter to 5.

The screenshot shows the configuration interface for the model 'flan-ul2-20b'. The 'Model parameters' section has 'Greedy' selected with a toggle switch. The 'Repetition penalty' is set to 1. The 'Stopping criteria' section shows two stop sequences: 'Question' and 'Problem'. The 'Min tokens' is set to 0 and the 'Max tokens' is set to 5. A blue button with a right arrow is visible on the left side of the interface.

- Step** 7  
Craft the Prompt text: paste this zero-shot prompt text into the freeform prompt editor in Prompt Lab.

1. 1
2. 2
3. 3
4. 4



5. 5
6. 6
7. 7
8. 8
9. 9
10. 10
11. 11
12. 12
13. 13

- 1.
2. **Classify this customer message into one of two classes: Question, Problem.**
3. **Class name: Question\***
- 4.
5. **\*Description: The customer is asking a technical question or a how-to question about our products or services.\***
- 6.
7. **\*Class name: Problem\***
- 8.
9. **\*Description: The customer is describing a problem they are having. They might say they are trying something, but it's not working. They might say they are getting an error or unexpected results.\***
- 10.
11. **\*Message: I'm having trouble registering for a new account.\***

Copied!

Prompt Lab

New (unsaved)

New prompt +

Save work ▾

Structured

Freeform

Model: flan-ul2-20b ▾

</>

🔧

↔

Classify this customer message into one of two classes: Question, Problem.

ⓘ

Class name: Question  
Description: The customer is asking a technical question or a how-to question about our products or services.

Class name: Problem  
Description: The customer is describing a problem they are having. They might say they are trying something, but it's not working. They might say they are getting an error or unexpected results.

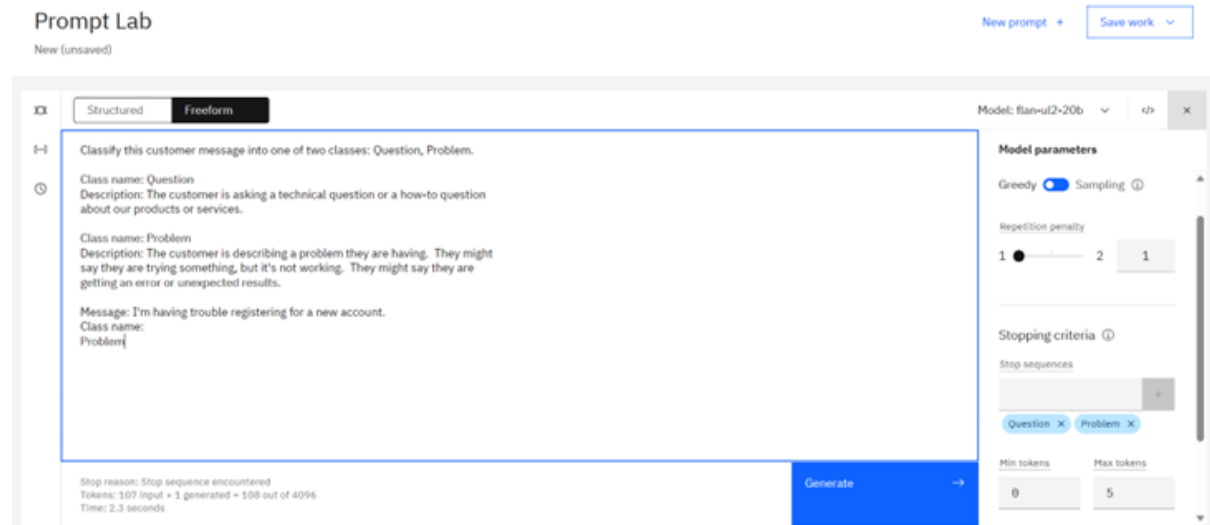
Message: I'm having trouble registering for a new account.  
Class name:

Generate →

### Step

Click **Generate** to see the results.

It will be able to classify the message to corresponding Class either Question or Problem.



## Exercise 3: Generate a numbered list

### Scenario

Say, that you want to generate a numbered list on a given theme using a few-shot prompt.

#### Step 1

Click on the **New Prompt** Button.

#### Step 2

Chose a model: `flan-ul2-20b` is a good choice.

#### Step 3

Set the decoding to **Sampling** and set the following parameters:

1. Temperature: 0.7
2. Top P: 1
3. Top K: 50

#### Step 4

Set the **Random seed** to 9045 to get a different output each time you click Generate. Specify a different value for the Random seed parameter or clear the parameter.

Model: flan-ul2-20b

</>

×

### Model parameters

#### Decoding

Greedy ☒ Sampling ⓘ

Temperature

0 ————— 2

Top P (nucleus sampling)

0 ————— 1

Top K

1 ————— 100

Random seed

Step  
Identify the **Stopping criteria**:

Model: flan-ul2-20b
</>
X

### Model parameters

Repetition penalty

1
2

### Stopping criteria ⓘ

Stop sequences

+

↵ ↵ X

Min tokens

Max tokens

Enter up to 6 sequences to stop output after the minimum number of tokens is reached.

To make sure the model stops generating text after one list, specify a stop sequence of two newline characters. To do that, click the **Stop sequence** text box, press the Enter key twice, then click Add sequence. Max tokens parameter to 50.

Step

6

Craft the prompt text: paste this few-shot prompt text into the freeform prompt editor in Prompt Lab.

1. \*What are 4 types of dog breed?
2. 1.Poodle
3. 2.Dalmatian

4. 3.Golden retriever
5. 4.Bulldog\*
- 6.
7. \*What are 3 ways to incorporate exercise into your day?
8. 1.Go for a walk at lunch
9. 2.Take the stairs instead of the elevator
10. 3.Park farther away from your destination\*
- 11.
12. \*What are 4 kinds of vegetable?
13. 1.Spinach
14. 2.Carrots
15. 3.Broccoli
16. 4.Cauliflower\*
- 17.
18. \*What are the 3 primary colors?
19. 1.Red
20. 2.Green
21. 3.Blue\*
- 22.
23. \*What are 3 ingredients that are good on pizza?''\*

Copied!

Prompt Lab

New (unsaved)

New prompt + Save work v

Model:flan-ul2-20b

Structured Freeform

4. Foodie  
2. Dalmatian  
3. Golden retriever  
4. Bulldog

What are 3 ways to incorporate exercise into your day?

1. Go for a walk at lunch  
2. Take the stairs instead of the elevator  
3. Park farther away from your destination

What are 4 kinds of vegetable?

1. Spinach  
2. Carrots  
3. Broccoli  
4. Cauliflower

What are the 3 primary colors?

1. Red  
2. Green  
3. Blue

What are 3 ingredients that are good on pizza?

Generate →

Model parameters

Repetition penalty

1 2 1

Stopping criteria ⓘ

Stop sequences

++ X

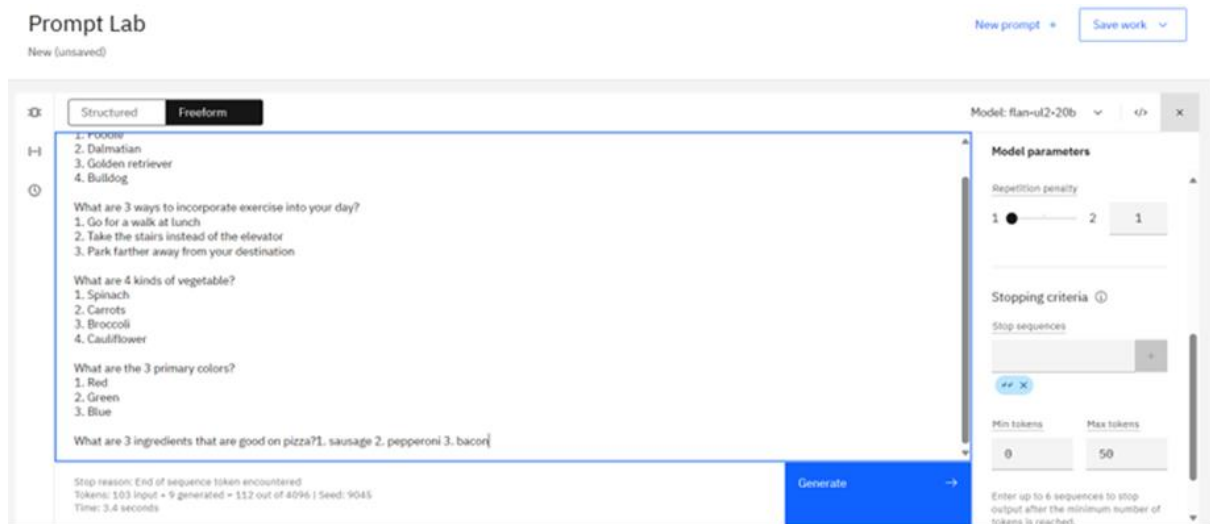
Min tokens Max tokens

0 50

Enter up to 6 sequences to stop output after the minimum number of tokens is reached.

## Step

Click **Generate** to see the result.



Three ingredients are suggested.

## Conclusion

The Prompt Lab is a versatile tool. Its AI models have helpful capabilities such as generating creative content, answering complex queries, assisting in language translation, summarizing content, extracting specific data from unstructured data, and even simulating specific roles like an English professor, as seen in this lab. It's a testament to the power and utility of modern AI.