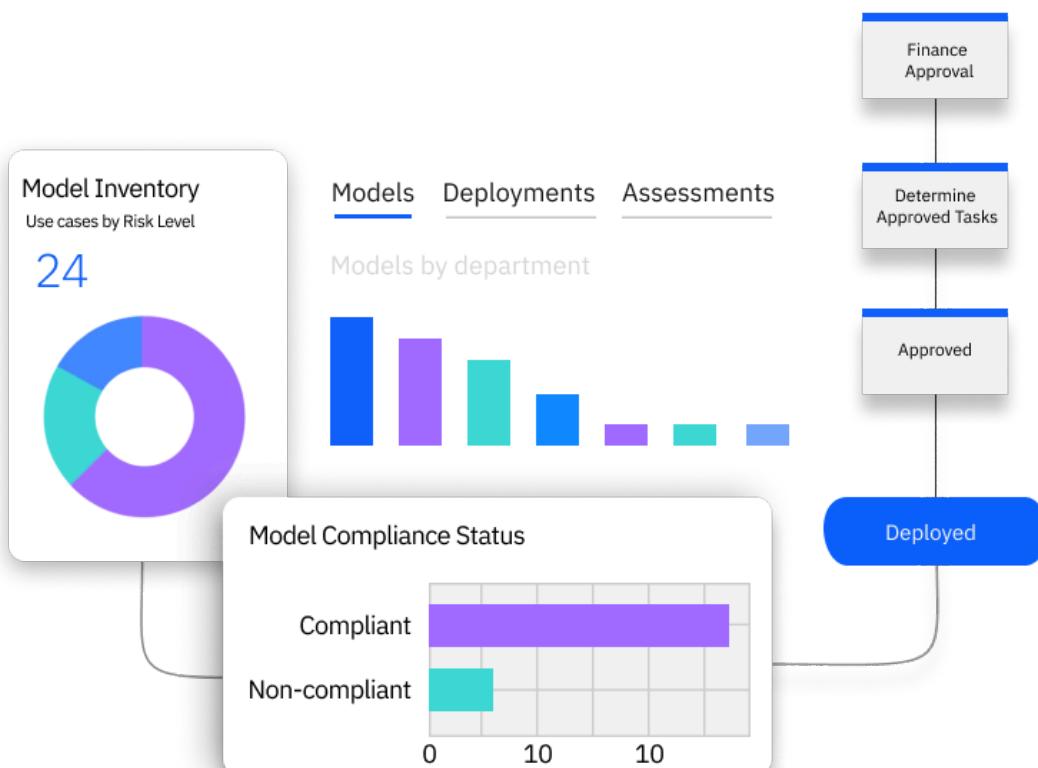


# watsonx.governance

## Governing Third Party Generative Models

Hands-on Lab Guide



Eric Martens  
emartens@us.ibm.com  
Information Developer, WW Technology

# Contents

1. Introduction
  1. About this hands-on lab
2. Signing into the environment and gathering credentials
3. Launch the governance service
4. Define a use case
5. Associate a project
6. Access the OpenAI model
7. Track the prompt
8. Evaluate the prompt
9. Review the evaluation
10. Conclusion
11. Troubleshooting
  1. No inventory / use cases disappearing
  2. Project imports, model evaluations or other IBM Cloud operations failing
12. Did not complete the lab?

# Governing third-party generative models

## Introduction

The pace of innovation in generative AI has only accelerated in the last few years, with multiple vendors regularly releasing new models with ever more advanced capabilities. In this environment, organizations looking to infuse AI in their day-to-day operations cannot afford to be tied to a specific vendor or set of development tools.

IBM watsonx.governance is a core component of watsonx, IBM's enterprise-ready AI and data platform that's designed to multiply the impact of AI across an enterprise. It offers tools for tracing, monitoring, and evaluating generative AI on any platform.

### 1. About this hands-on lab

The main objective of this lab is to experiment with the watsonx.governance tools for monitoring 3rd party AI models. Users will interact with prompts via Jupyter notebook, and see the kinds of evaluations and metrics available.

## Signing into the environment and gathering credentials

As part of your lab reservation, you have received a [login URL](#), as well as a username and password.

 It is **HIGHLY** recommended to use a new incognito or private browser window for this lab to prevent any cache or cookie conflicts. You may also clear your browser's cache. If you run into any issues in the lab with access to projects or services, they can usually be fixed by clearing your cache or opening a new private browser window and navigating back to the lab environment.

1. In your browser, navigate to the [login URL](#), enter your provided credentials (A) and click on the [Sign in](#) button (B). You may need to wait a few seconds for the account to be fully configured.



Once the configuration has completed, your first task will be to create an API key that will allow you to authenticate with the watsonx.ai and watsonx.governance services.

2. Click on the Manage menu (A) to open it, then click on the Access (IAM) menu item (B). The IBM Cloud Identity and Access Management screen opens.

The screenshot shows the IBM Cloud dashboard. At the top, there's a search bar and a 'Manage' button with a dropdown arrow. A red circle labeled 'A' highlights the 'Manage' button. The dropdown menu is open, showing options like 'Account', 'Billing and usage', 'Catalogs', 'Enterprise', 'Security and access', and 'Access (IAM)'. A red circle labeled 'B' highlights the 'Access (IAM)' option. To the right of the menu, there's a 'Create resource' button and a 'Select an option' dropdown.

3. From the Manage identities section of the menu on the left, click on the API keys entry (A). Click on the Create button (B). The Create IBM Cloud API key modal window opens.

The screenshot shows the 'API keys' page under the 'Manage identities' section of the left sidebar. The 'API keys' entry is highlighted with a red circle labeled 'A'. On the right, there's a table of API keys with a 'Create +' button at the top right, which is also highlighted with a red circle labeled 'B'.

4. Give your key a name (A) and an optional description (B). Click on the Create button (C) to create the key.

The screenshot shows the 'Create IBM Cloud API key' modal window. In the 'Name' field, 'techxchange lab' is entered, with a red circle labeled 'A' highlighting it. In the 'Description (optional)' field, 'Complete lab activities for governing third-party models' is typed, with a red circle labeled 'B' highlighting it. At the bottom right of the modal, there's a 'Create' button with a red circle labeled 'C' highlighting it.

5. Click on the copy button (A) to copy the key to your clipboard, then click on the close button (B) to close the modal window. Note that you will be unable to access the key after you close the modal window, so be sure to copy it to your machine's clipboard before closing it. If you do not copy your key, you can delete it and create a new one.

API key successfully created

Copy the API key or click download to save it. You won't be able to see this API key again, so you can't retrieve it later. The API key is no longer displayed after 295 seconds.

**API key**

**Status**

**Enabled**

**Yes**

**Copy** **Download**

6. Paste the API key into a text editor or other application on your machine to save it for later use.

You can now proceed with the watsonx.governance service

## Launch the governance service

1. Click on the **Navigation menu** button (A) to expand it, then click on the **Resource list** button (B). The resource table opens.

IBM Cloud

Search resources and products...

Catalog Manage 3062701 - itz-saas-058

Dashboard

Projects

Resource list

Containers

Databases

Infrastructure

Observability

Platform Automation

Security

API Management

API keys

Create, view, and work with API keys that you have access to manage. IBM Cloud API keys are associated with a user's identity and can be used to access cloud platform and classic infrastructure APIs, depending on the access that is assigned to the user. The following table displays a list of API keys created in this account. [Learn more](#).

Looking for more options to manage API Keys? Try [IBM Cloud® Secrets Manager](#) for creating and leasing API keys dynamically and storing them securely in your own dedicated instance.

Unused or overly permissive API keys increase the risk of unauthorized access. Regularly review the [Inactive identities report](#), rotate keys, and apply only the minimum required permissions.

API keys associated with a user's identity have the same access that the user is assigned across all accounts. To update the access for an API key, assign or remove access for the user.

View My IBM Cloud API keys

2. From the resource list table, click on the **AI / Machine Learning** section (A) to expand it. Locate the **watsonx.governance** service from the list (B) and click on the linked name to open it.

**⚠️** If you have more than one watsonx.governance service, choose the service in **Dallas (us-south)**.

Resource list

Create resource +

Name	Group	Location	Product	Status	Tags
gov-550000w6t0	eid-68ca2705f262a7ad7bad874b	Dallas (us-south)	watsonx.governance	Active	eid:68c... +1
wml-550000w6t0	eid-68ca2705f262a7ad7bad874b	Dallas (us-south)	watsonx.ai Runtime	Active	eid:68c... +1
ws-550000w6t0	eid-68ca2705f262a7ad7bad874b	Dallas (us-south)	watsonx.ai Studio	Active	eid:68c... +1

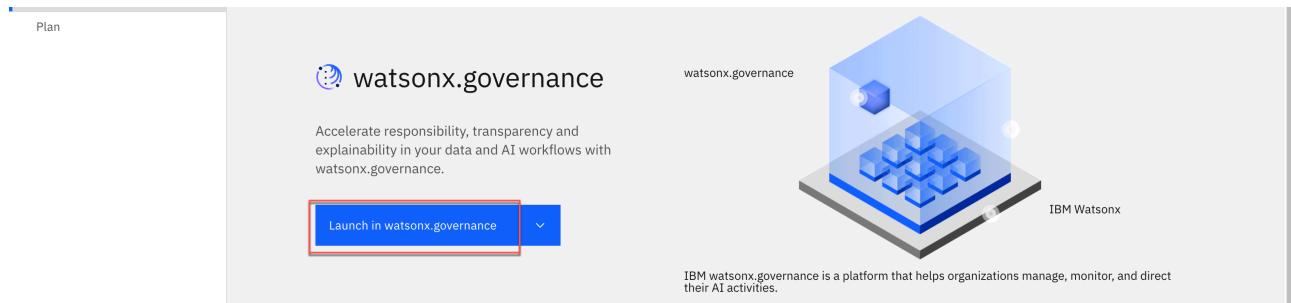
AI / Machine Learning (3)

Analytics (1+)

Application development (0)

Asset management (0)

3. Click on the **Launch in watsonx.governance** button to launch the service in a new browser tab.



4. Check the box in the [Welcome to watsonx](#) popup to agree to the terms and conditions. Then click the [close](#) button to close the popup.

## Define a use case

In watsonx.governance, a [use case](#) is defined as a problem or opportunity that an organization would like to address using AI or machine learning. Example use cases include predicting credit risk, evaluating loan applications, answering customer service inquiries, and more. Use cases can be associated with multiple machine learning models or generative AI prompts at various points in the lifecycle, including both production models and candidate models in development. Use cases represent a foundational building block for AI governance.

In this lab, you are attempting to use a generative AI prompt to summarize candidate resumes to assist in the hiring process.

1. Click on the [Navigation menu](#) (A) to open the menu. Then click on the [AI governance](#) menu item (B) to expand it. Finally, click on the [AI use cases](#) menu item (C). The use cases screen opens. Close any popup windows that may appear.

2. Click on the [New AI use case](#) button. The [New AI use case](#) screen opens.

⚠️ This lab is a shared environment that is being accessed by other users. You may see their use case details listed on the screen.

3. Give your use case a unique **name** (A) and optional description (B). Click on the **Risk level** dropdown and select a risk level (C). Note that a full risk assessment is beyond the scope of this lab, but risk identification, assessment, and management is a core capability of watsonx.governance.

The screenshot shows the 'New AI use case' interface. On the left, there's a sidebar with 'AI' and 'Narrative' sections. The main area has a title 'New AI use case' and a subtitle 'Create a use case to define a business problem, request a model, and specify details such as risk level and status.' It contains several input fields:

- Name (Required):** 'Resume summarization - emartens' (highlighted with a red box and labeled A).
- Description:** 'Summarize applicant resumes.' (highlighted with a red box and labeled B).
- Risk level:** 'Medium' (highlighted with a red box and labeled C).
- Inventory (Required):** 'Default Inventory (itz-saas-058)'.

On the right side, there are sections for 'Owner' (with a user icon and email 'student0@techzone.ibm.com'), 'Status' (set to 'Draft'), and 'Tags' (with a search bar and a note 'Add tags to this AI use case').

4. Click on the **Create** button in the lower right to create the use case. The **New AI use case** window closes. When your use case is created, its details will appear on the screen.

## Associate a project

Watsonx projects are collaborative environments where data engineers, data scientists, AI engineers, and subject-matter experts can work together, sharing code, data, and other assets to solve machine learning and AI problems. Projects can be associated with use cases, allowing those assets to be tracked and viewed as part of the model lifecycle. In this section of the lab, you will create a project that will allow you to work with and evaluate an OpenAI model hosted on Microsoft Azure as part of your use case.

**⚠ THESE EVALUATIONS ARE NOT INTENDED TO SHOW THE RELATIVE STRENGTHS OF THE OPENAI OR AZURE PLATFORMS.**

1. Scroll down to the **Associated workspaces** section of your use case. Note that three phases of the model lifecycle are listed here: **Development**, **Validation**, and **Operation**.
2. In the **Development** section, click on the **Associate workspaces** button. The **Associate workspaces** window opens.

The screenshot shows the 'Associated workspaces' window. On the left, there are three sections: 'Development' (with a note about experimenting with ML development), 'Validation' (with a note about evaluating models in pre-production), and 'Operation' (with a note about using deployment spaces and prompt templates). Each section has a 'Associate workspaces' button. To the right, there's a circular diagram with 'Workspaces for Develop' at the top, 'AI use case' in the center, 'Workspaces for Operate' at the bottom-left, and 'Workspaces for Validate' at the bottom-right. On the far right, the detailed properties of the AI use case are listed:

<b>Name</b>	Resume summarization - emartens
<b>Description</b>	Summarize applicant resumes.
<b>Owners</b>	(NN)
<b>Status</b>	Draft(notset notset, Sep 26, 2025)
<b>Risk level</b>	Medium
<b>Inventory</b>	Default Inventory
<b>Tags</b>	Add tags to this AI use case.

3. Scroll down to the **Projects** section of the window and click on the **New project** button. The **Create a project** window opens.

Projects

A project can be associated with only one lifecycle phase for an AI use case.

Find projects

Name	Created	Your role	Tracked AI assets	Associates

Projects will display here with information about associations with use cases.

4. Give your project a unique **name** (A). Click on the **Target Cloud Object Storage Instance** dropdown and select an object storage instance from the list (B).

Create a project

Start with a new, blank project or select from where to import an existing project.

Define details

Name Resume summarization - emartens A

Description (optional)  
What's the purpose of this project?

Tags (optional)  
Add tags

Define storage

Select storage service cos-550000w6t0 B

Project includes integration with [Cloud Object Storage](#) for storing project assets.

5. Click on the **Create** button to create the project. The **Create a project** window closes, and the new project is now listed as a workspace.
6. Click on the **Save** button to save your changes to the use case and return to the use case details screen.

## Access the OpenAI model

In the next step, you will use a Jupyter notebook in your project to access the OpenAI model.

1. In the **Development** tile of the **Associated workspaces** section of the use case details, click on the **open** icon for your project. The project opens in a new browser tab.

Associated workspaces

Associate your AI use case with the workspaces in order to organise them under the same business problem.

Development open

Validation

In this phase your validators can evaluate models in pre-production deployment spaces and prompt templates in projects.

Associate workspaces +

Workspaces for Develop

AI use case

Workspaces for Operate

Workspaces for Validate

Name	Resume summarization - emartens
Description	Summarize applicant resumes.
Owners	NN
Status	Draft(notset notset, Sep 26, 2025)
Risk level	Medium
Inventory	Default Inventory
Tags	(+)

2. Click on the **Assets** tab (A) to switch to it. Click on the **New asset** button (B).

The screenshot shows the IBM Watsonx interface. At the top, there's a navigation bar with 'IBM Watsonx' on the left and various project and account details on the right. Below the navigation bar, there are tabs: 'Overview' (highlighted), 'Assets' (highlighted with a red box and circled A), 'Jobs', and 'Manage'. On the left side, there's a sidebar with sections like 'Asset types' and a note about organizing assets by type. On the right, there's a main area titled 'All assets' with a search bar and a 'New asset' button at the top right.

3. Take a moment to review some of the options for assets that can be added to projects. Watsonx projects are highly flexible, and can be used to collaborate and manage access to a huge variety of tools, prompts, data, data connections, and more. When you are finished, locate and click on the **Work with data and models in Python...** tile in the **Work with models** section.

This screenshot shows a modal dialog titled 'What do you want to do?'. It asks to select a task based on a goal. On the left, there's a sidebar with 'Assets' selected. The main area has a search bar and several tiles. One tile, 'Work with data and models in Python or R notebooks', is highlighted with a red box and circled B. Other tiles include 'Evaluate and compare AI assets', 'Solve optimization problems', 'with Evaluation Studio', 'with Jupyter notebook editor', and 'with Decision Optimization'.

4. Click on the **URL** option (A) from the list on the left. Give your notebook a **name** (B). Copy and paste the following URL into the **Notebook URL** field (C):

This screenshot shows the configuration dialog for creating a new notebook asset. On the left, there's a sidebar with options: '+ New', 'Sample', 'Local file', and 'URL' (highlighted with a red box and circled A). The main area is divided into three sections: 'Define details' (with 'Name' field containing 'Create prompt template' and circled B), 'Define configuration' (with 'Select runtime' dropdown set to 'Runtime 24.1 on Python 3.11 XS (2 vCPU 8 GB RAM)' and descriptive text), and 'Notebook URL' (with a text input field containing the URL 'https://github.com/ericmartens/governing-third-party/raw/refs/heads/main/create\_prompt\_template\_saas.ipynb' and circled C).

5. Click on the **Create** button to create your notebook from the notebook in github.

The notebook consists of two types of cells: markdown cells, which contain explanatory text, and code cells like the one below:

```
[ ]: import requests
import urllib3, json # noqa: E401
urllib3.disable_warnings()

def generate_access_token():
    headers={}
    headers["Content-Type"] = "application/json"
    headers["Accept"] = "application/json"
    data = {
        "username":CPD_USERNAME,
        "api_key":CPD_API_KEY
    }
    data = json.dumps(data).encode("utf-8")
    url = CPD_URL + "/icp4d-api/v1/authorize"
    url = url.replace("//icp4d", "/icp4d")
    response = requests.post(url=url, data=data, headers=headers, verify=False)
    response.raise_for_status()
    json_data = response.json()
    iam_access_token = json_data['token']
    print("Access token generated successfully!")
    return iam_access_token

iam_access_token = generate_access_token()
```

Notebooks are designed to be run one cell at a time from top to bottom. You can click on a cell to highlight it and make it active. You can run an active cell by pressing the **Shift + Enter** keyboard shortcut, or by clicking on the run button in the task bar:



Running the active cell will execute any code in the cell, and will make the following cell active. If the cell contains only markdown text, no action will be taken and the next cell will become active. As a cell runs, an asterisk (\*) will appear in the brackets to the left of the cell. When it completes, the asterisk will be replaced by a number. You can run active cells while code is still executing; they will execute in the order you ran them when the previous cell is finished, unless interrupted by an error.

6. Scroll down to the second code cell that contains the variables for your API key that you created earlier in the lab and the Azure credentials provided by your lab instructor. Enter those variables in the cell, making sure to leave the values enclosed in quotation marks as shown in the screen below.

```
*[5]: import os
from rich import print
from IPython.display import display, Markdown

CLOUD_API_KEY = "ogK6CY"                                     'wFpNH8"
AZURE_OPENAI_ENDPOINT = "https://azureml-openai-americas-1.openai.azure.com/" 
AZURE_OPENAI_DEPLOYMENT_NAME = "tz-gpt-35-turbo-americas-1"
AZURE_CLIENT_ID = "98be066"                                  e3c5"
AZURE_CLIENT_SECRET = "84E8I"                                 ObdM"
AZURE_TENANT_ID = "4e7";                                    61b7"

DATAPLATFORM_URL = os.environ.get('RUNTIME_ENV_APXS_URL')
IAM_URL = "https://iam.cloud.ibm.com"
PROJECT_ID = os.environ.get('PROJECT_ID', '<YOUR_PROJECT_ID>')
print(f'Your project id is {(PROJECT_ID)} and your API URL is {CPD_URL}')
```

7. Run through the cells in the notebook one at a time, taking a moment to read and understand the text above each cell. The notebook will sign in to the Watson services using your API key. Next, it will define a default prompt that uses ChatGPT to summarize resumes, and connect to the ChatGPT model running in Microsoft Azure. It will then load sample resumes from GitHub to test the output of the prompt, exporting that output as a data asset in the project, which will be used in later steps to evaluate the model. Finally, it saves the prompt with the associated credentials as a project asset known as a *detached prompt template*.

8. When you have finished running all the cells in the notebook, click on the **breadcrumb link** in the upper right to return to your project.

IBM watsonx

Projects / [Resume summarization - emartens](#) / Create prompt template

```
eval_df = data.drop(['extraction', 'Resume_without_profile'], axis=1)
eval_df.to_csv("resume_summarization_eval_data.csv", index=False)
res = wslib.upload_file("resume_summarization_eval_data.csv", overwrite=True)
```

Create the detached prompt template

Create a detached prompt template in your project for the summarization task that references the Azure OpenAI model.

```
In [ ]: from ibm_aigov_facts_client import (
    AIGovFactsClient, CloudPakForDataConfig,
    DetachedPromptTemplate, PromptTemplate
)
from ibm_aigov_facts_client.utils.enums import Task
```

9. In order to evaluate your prompt, you will need to use the watsonx.ai service. Your account reservation has access to this service, but you will need to associate it with your project. Click on the [Manage tab](#) (A). Click on the [Services & integrations](#) item (B). Click on the [Associate service](#) button (C).

The screenshot shows the IBM WatsonX interface with the 'Manage' tab highlighted (A). Under the 'Services & integrations' section (B), the 'Associate service' button (C) is highlighted with a red circle.

10. Check the box to the left of the [watsonx.ai Runtime](#) service.

The screenshot shows the 'Associate service' dialog. A service named 'wml-550000w6t0' (Type: watsonx.ai Runtime) is selected (indicated by a checked checkbox). The 'Associate' button is highlighted with a red circle.

11. Click on the [Associate](#) button to finalize the association.

## Track the prompt

At this point, you should see three assets listed on the [Assets](#) tab of your project: the Jupyter notebook, a CSV file containing evaluation data, and the detached prompt template for the ChatGPT model. You can now start to track the model in your use case.

1. Click on the [Assets](#) tab (A), then click on the link for the detached prompt template to open the AI Factsheet for the prompt (B). A factsheet captures essential model details that are required for AI model governance and compliance and tracks the following metadata across the model development lifecycle. A popup window will open, inviting you to take a brief tour of the AI Factsheet. You can take the tour, or choose to skip it.

The screenshot shows the WatsonX Assets interface. The top navigation bar includes tabs for Overview, Assets (highlighted with a red box and labeled 'A'), Jobs, and Manage. Below the navigation is a search bar with the placeholder 'Find assets'. On the left, a sidebar displays '3 assets' and categories: Asset types (Data, Notebooks, Prompts), Foundation model, Prompt template, Prompt parameters, and Development (Resume summarization, Additional details, Attachments). The main area is titled 'All assets' and lists three items:

Name	Last modified
Create prompt template Notebook from URL	3 days ago Modified by you
<a href="#">Detached prompt for Azure OpenAI GPT-3.5-turbo</a> Detached prompt template	3 days ago Modified by Service
resume_summarization_eval_data.csv CSV	3 days ago Modified by you

2. Take a moment to review the information that has already been collected on the factsheet, and recall that much of it (including the model provider information) was defined by the Jupyter notebook. When you are finished, scroll back to the top of the factsheet and click on the [Track in AI use case](#) button. This will allow you to configure tracking for the prompt in the use case you created earlier. Recall that you already associated your watsonx project with the use case's [Development](#) lifecycle phase.

The screenshot shows the WatsonX Governance interface for a tracked prompt template. The left sidebar shows the same asset categories as the Assets screen. The main area is titled 'Governance' and contains the following message:

This prompt template is not tracked.  
To track a prompt template, add it to an AI use case. Tracking captures details about the asset for governance purposes.  
**Important:** Once you start tracking a prompt template in a use case, you can no longer edit it. Wait until the prompt template is stable to start tracking.

At the bottom, there is a blue button labeled 'Track in AI use case' with a red box around it, and a link 'Learn more'.

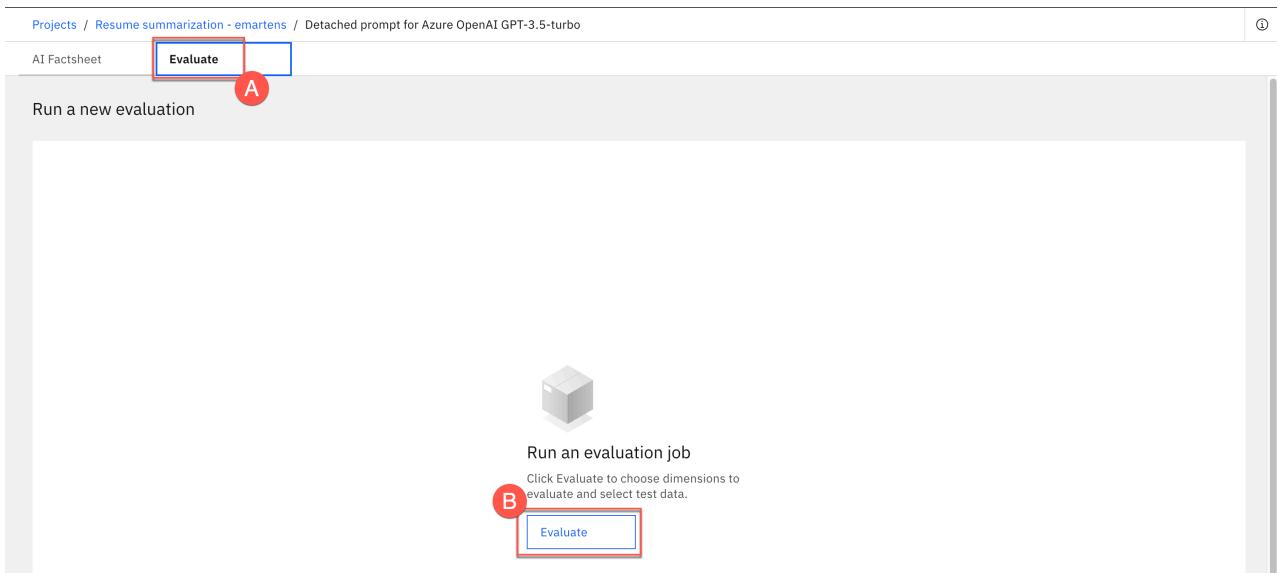
3. Click on the [Next](#) button to accept the default model tracking approach.
4. You can now assign a version number to the tracked prompt by clicking one of the tiles, or using the text entry field beneath them to manually assign a number. Choose one of the options, then click on the [Next](#) button to continue.
5. On the [Review](#) screen, note that once you enable tracking for a prompt, you will no longer be able to make any changes to it, as any alterations would invalidate performance metrics. Any additional iterations or enhancements to the prompt would be tracked as a different prompt template within the same use case.

Click on the [Track asset](#) button to finalize asset tracking.

## Evaluate the prompt

Now that the prompt is tracked as part of the use case, any performance metrics gathered will automatically be stored in the factsheet, providing a complete picture of model lineage and performance.

1. To run an evaluation, click on the [Evaluate](#) tab (A) to open it, then click on the [Evaluate](#) button (B). The [Evaluate prompt template](#) window opens.



2. Two different evaluation dimensions are available. The prompt will automatically be scored for model health metrics; however, you can choose whether you wish to also calculate generative AI quality metrics. For those metrics, you can set acceptable thresholds for the prompt. Click on the [Advanced settings](#) button. The [Configure evaluations](#) window opens.

Evaluate prompt template

Choose the evaluation dimensions and select the test data. [Learn more](#)

<input checked="" type="radio"/> Select dimensions	Select dimensions to evaluate These dimensions are based on the prompt template task type. <a href="#">Learn more</a>
<input type="radio"/> Select test data	Language For input and output, select the associated language. English
<input type="radio"/> Map variables	
<input type="radio"/> Review and evaluate	

[Advanced settings](#)

<input checked="" type="checkbox"/> Dimension	Description
<input checked="" type="checkbox"/> Generative AI Quality	The Generative AI Quality monitor calculates a variety of metrics based on prompt template task type. Some metrics compare model output to the reference output you provide. Other metrics analyze model input and output and do not require reference output.
<input checked="" type="checkbox"/> Model health	The model health monitor provides an overview and helps to understand how your model deployment is performing with the incoming transactions. This monitor evaluates and computes metrics such as scoring requests count, records count, latency and throughput, etc.

3. Click on the [edit icon](#) in the [Settings](#) tile.

Configure evaluations

Evaluations

- Generative AI Quality
- Model health

Generative AI Quality

Description

The Generative AI Quality monitor calculates a variety of metrics based on prompt template task type. Some metrics compare model output to the reference output you provide. Other metrics analyze model input and output and do not require reference output.

Settings

Task type: Text summarization

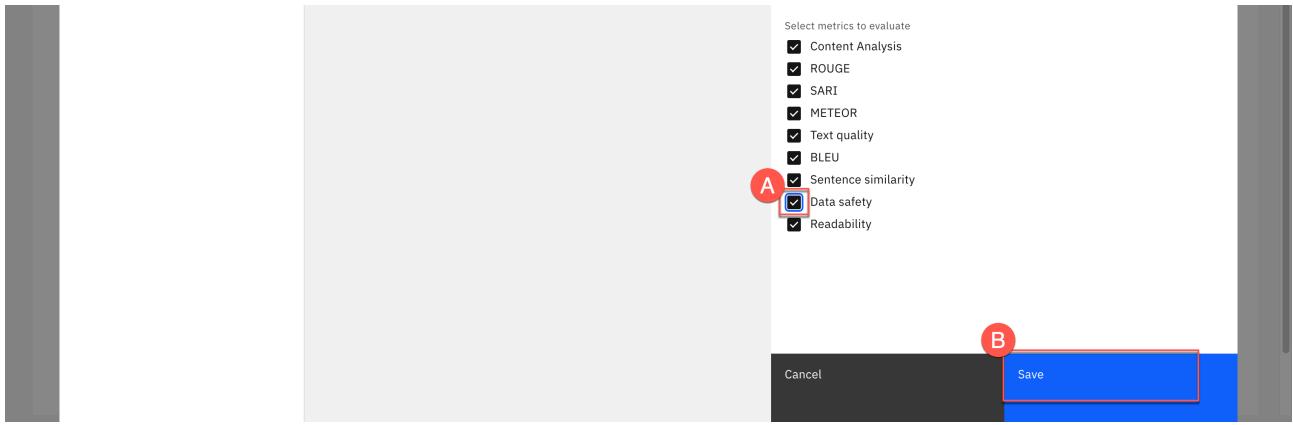
Selected metrics: Content Analysis, ROUGE, SARI, METEOR, Text quality, BLEU, Sentence similarity, Readability

Content Analysis

Ngrams (Abstractness)  
1

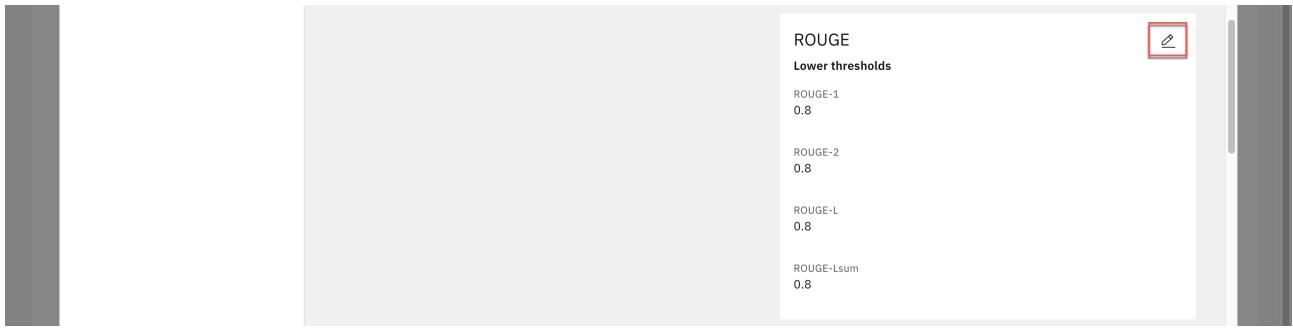
Ngrams (Repetitiveness)

4. Note the wide range of metrics available. Check the box to the left of the [Data safety](#) item (A) to turn on evaluations for those items, then click on the [Save](#) button (B). The [Configure settings](#) window closes.



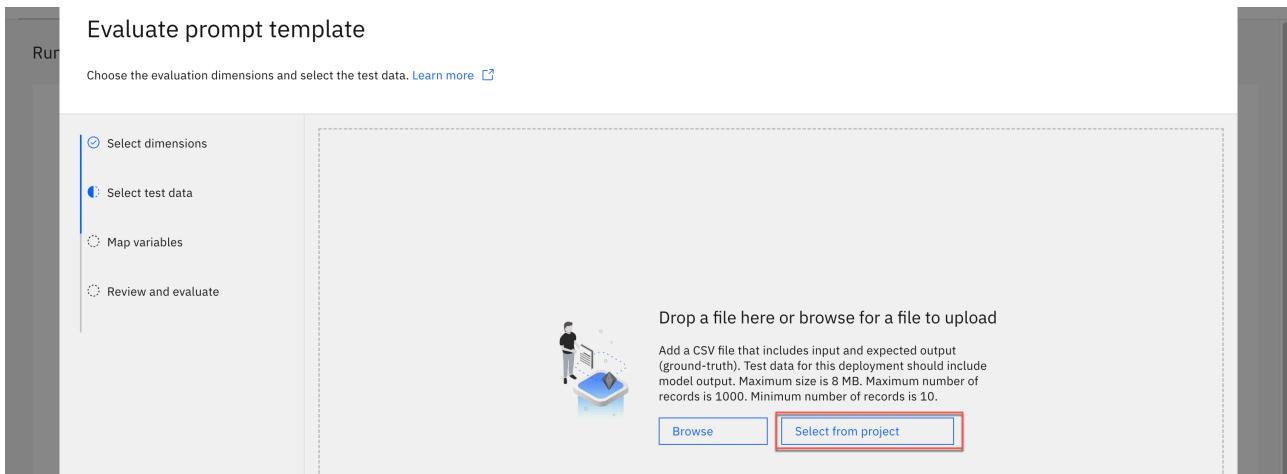
Each group of metrics (Content Analysis, ROUGE, SARI, etc.) can be individually configured using the tiles on the right of the screen. Because you turned on the data safety monitors in the previous step, a [Data safety](#) tile now appears on this screen.

5. Scroll down to the **ROUGE** tile and click on the [edit icon](#). The information panel changes to show an overview of the metric, along with alert thresholds on the left.



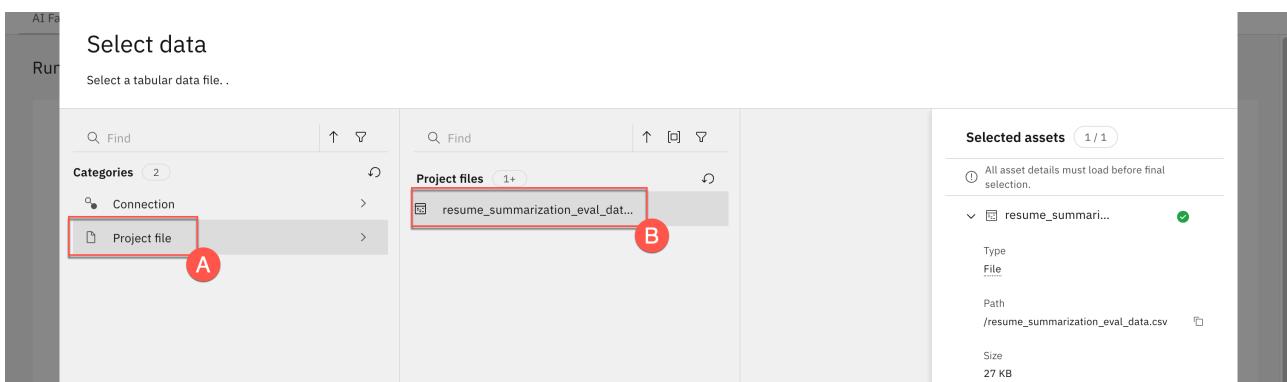
Note that you can click on the [information icon](#) for each metric to get an expanded description. Take a moment to learn about how ROUGE is calculated and what each of the different calculations mean. You may also alter the alert thresholds; shortly, you will run a model evaluation. If the calculated metric score falls below the set alert threshold, the test result will register as a failure.

6. Click on the [Save](#) button to save any changes you have made and close the **ROUGE** configuration panel.
7. Click on the [Save](#) button again to save your changes, close the [Configure evaluations](#) panel, and return to the [Evaluate prompt template](#) screen.
8. Click on the [Next](#) button. The [Select test data](#) step opens. From here, you can specify the evaluation data for your prompt.
9. Recall that the Jupyter notebook exported evaluation data to your watsonx project. Click on the [Select from project](#) button. The [Select data](#) window opens.

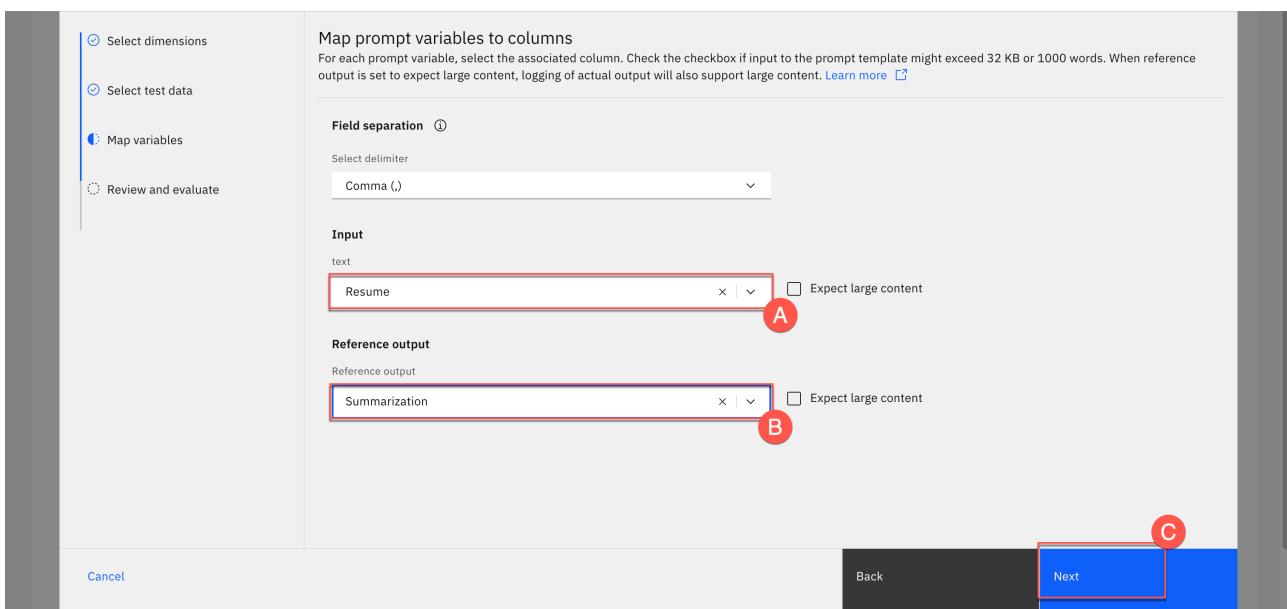


⚠️ If you receive a [Failed to load project](#) error message, you can click on the [Cancel](#) button to close the [Select data](#) window, then click on the [Select from project](#) button again. If you continue to receive the message, you may need to refresh your browser or log in using a new private / incognito window.

10. Click on the [Project file](#) option (A). From the list of project files, click on the [resume\\_summarization\\_eval\\_data](#) file (B) to select it.



11. Click on the [Select](#) button to load the CSV file and progress to the [Map variables](#) section.
12. Click on the [Input](#) dropdown and select [Resume](#) (A). Click on the [Reference output](#) dropdown and select [Summarization](#). Click on the [Next](#) button (C) to advance to the [Review and evaluate](#) step.



13. Check the box to allow the service to use task credentials.

The screenshot shows the 'Evaluate prompt template' interface. On the left, there's a sidebar with 'AI Pa' and 'Run'. The main area has a title 'Evaluate prompt template' and a sub-instruction 'Choose the evaluation dimensions and select the test data. [Learn more](#)'. A vertical list on the left lists four steps: 'Select dimensions' (radio button selected), 'Select test data' (radio button selected), 'Map variables' (radio button selected), and 'Review and evaluate'. To the right, under 'Review', are details: Task: Text summarization, Test data: resume\_summarization\_eval\_data.csv, Language: English, and Evaluations: Generative AI Quality Model health. Below this is a section titled 'Task credentials' with a note 'Allow service to create, use, and store your credentials via task credentials' and a checked checkbox.

14. Click on the **Evaluate** button to start the evaluation. Note that it may take up to ten minutes to complete.

## Review the evaluation

Once the evaluation completes, the results will display on the screen. Note that your evaluations may differ from the screenshots below, but it is very likely that the prompt has failed several of the quality tests.

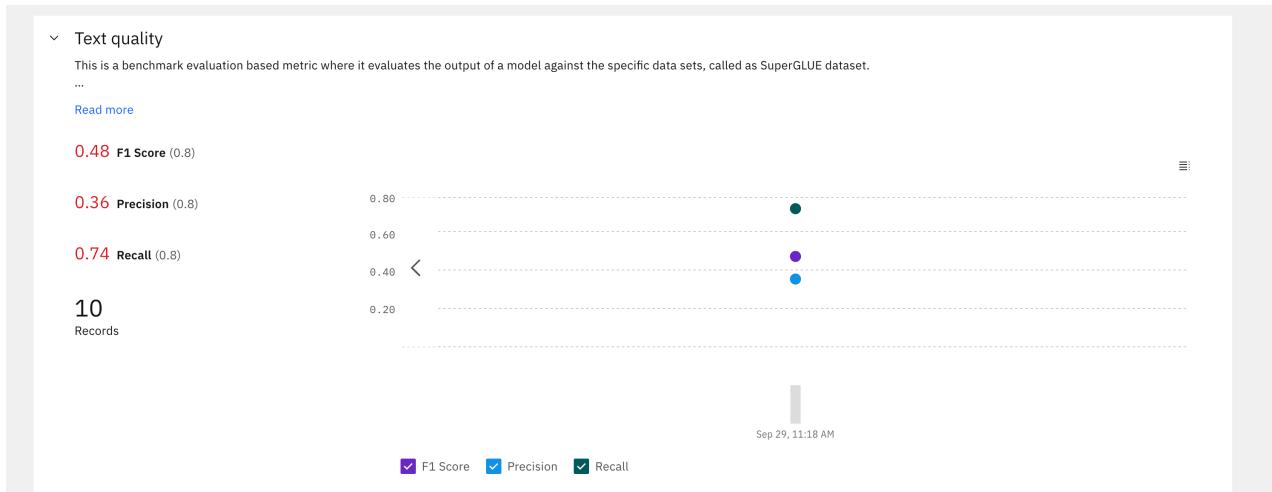
1. Click on the **View metrics** button. The expanded metrics view opens.

The screenshot shows the 'Generative AI Quality - Text summarization' metrics view. At the top, there's a red box around the 'Alerts triggered' button, which shows '16'. Below it is a table with columns 'Metric', 'Score', and 'Violation'. The table has two rows: 'SARI' with a score of 45.63 and a violation of 34.37, and 'METEOR' with a score of 0.59 and a violation of 0.21. There are expandable sections for 'Content Analysis' and 'ROUGE'.

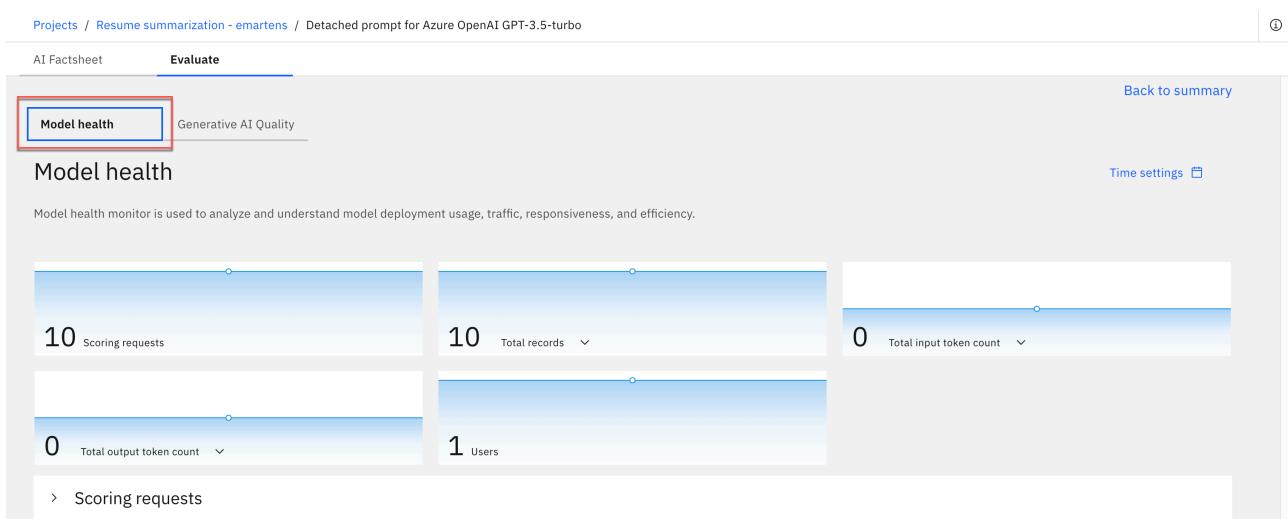
Metric	Score	Violation
SARI	45.63	34.37
METEOR	0.59	0.21

The detailed view for quality shows the different metrics over time; as more evaluations are performed, these graphs will update with the additional data points. Note that clicking on the **Time settings** link allows you to adjust the time window for the evaluations you would like to see.

2. Take some time to review the collected metrics. You can click on each panel to expand it and view a chart. Hovering your mouse over the chart will provide additional information.



- When you are finished, scroll back to the top of the window and click on the **Model health** tab to open it.



- Take a moment to review this tab, which contains historical data for health metrics such as token counts, number of users, and more. This information can be vital for an organization's infrastructure and engineering teams ensuring that the models are responding to application and user requests and keeping compute costs to acceptable levels.

## Conclusion

Congratulations, you have finished the [Evaluate and Monitor 3rd party Generative AI Models with watsonx.governance](#) lab. In this lab, you saw how you can organize and govern AI tasks with use cases. You created a watsonx project to collaborate with data scientists, data engineers, and AI engineers on AI and machine learning initiatives. You associated that project with your use case, then used a Jupyter notebook to interact with an OpenAI model running on Microsoft Azure. Finally, you enabled tracking for the model, configured thresholds for model metrics, and ran an evaluation using test data.

## Troubleshooting

Experiencing issues with the lab? Try these steps.

## 1. No inventory / use cases disappearing

This is by far the most common issue encountered when trying to complete the lab. To fix the issue, simply log out and log back in, or log in again using a new private/incognito browser window. In some rare cases, you may need to clear your entire browser cache, or use a completely different browser. However, in nearly all instances, logging in with a new private/incognito browser window will resolve the problem.

## 2. Project imports, model evaluations or other IBM Cloud operations failing

Due to the nature of Software-as-a-Service, you may experience intermittent issues such as:

- Project imports failing with unknown errors (not policy capacity limit errors; see the following section for that)
- Project assets importing successfully, but then not appearing in the asset lists
- Model evaluations timing out and failing
- Cloud object storage instances not appearing

In the vast majority of cases, these issues can be corrected by simply re-trying the task that failed. In some cases, model evaluations can take three or four tries to succeed. If IBM Cloud is under heavy load, waiting an hour or so and re-trying the task may succeed. The development teams responsible for the services have been made aware of the issues lab participants have faced with service outages, and we are working with them to resolve those issues and improve service stability.

## Did not complete the lab?

The lab environments will remain available for the remainder of the day. You may continue working on the lab as necessary.