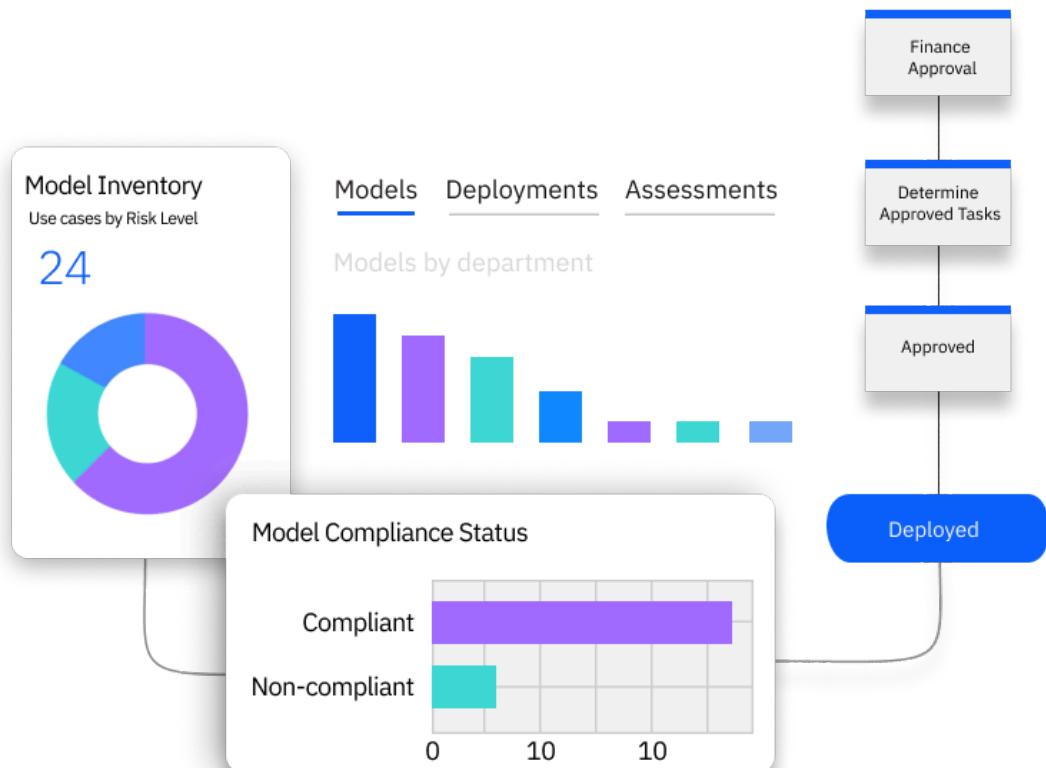


Monitor and Govern AI Models

watsonx.governance

Hands-on Lab Guide



Eric Martens
emartens@ibm.com
WW Learning Content Development

Contents

- 1. Introduction
- 2. Getting help - PLEASE READ
 - 1. Use case
 - 2. Environment
- 3. Reserving a TechZone account
- 4. Configuring a TechZone account
 - 1. Configure the monitoring service
 - 2. Identify a model inventory
- 5. Governing predictive models
 - 1. Set up a watsonx project
 - 2. Define a model use case
 - 3. Associate a project with the use case
 - 4. Associate a space with the use case
 - 5. Track the model
 - 6. Deploy the model
 - 7. Evaluate the model
 - 8. Explain a prediction
 - 9. View the updated lifecycle
 - 10. Next steps
- 6. Governing generative models
 - 1. Define an AI use case
 - 2. Create a new deployment space
 - 3. Set up a watsonx.ai project
 - 4. Associate use case workspaces
 - 5. Track the foundation model
 - 6. Deploy the foundation model
 - 7. Evaluate the foundation model
 - 8. View the updated lifecycle
- 7. Conclusion
- 8. Troubleshooting
 - 1. No inventory / use cases disappearing
 - 2. No access to services / services not appearing
 - 3. Compute capacity limits exceeded
 - 4. Project imports, model evaluations or other IBM Cloud operations failing
 - 5. Deployment monitoring or policy capacity limits exceeded
 - 6. Internal server error when accessing the monitoring service

Introduction

Welcome to the watsonx.governance Level 3 technical lab. In this lab, you will explore the governance capabilities of IBM's solution for both generative and predictive models. Many people going through this lab may already be familiar with governing predictive models; for this reason, the lab has been split into two wholly independent sections. If you are comfortable demoing predictive model governance for clients, you can begin by reading the introduction and setting up the use case inventory. When instructed, you can skip the predictive portion of the lab and focus entirely on the new generative AI governance capabilities.

Getting help - PLEASE READ

This is an extremely lengthy, highly technical lab that touches on multiple products and environments that are all under active development. Every effort has been made to address possible causes and issues in the instructions themselves; however, it is not uncommon for problems to arise, error messages to appear, or screens to sometimes differ from the lab instructions.

 PLEASE refer to the [Troubleshooting](#) section of the lab first to see if your problem is addressed there. That section will be continually updated to respond to the most frequent issues encountered in running the lab.

If your issue is not addressed, [CONTACT THE AUTHOR VIA SLACK](#) if at all possible. IBMers can reach Eric Martens [via Slack](#). Business partners can reach out via [email](#).

Leaving comments on the YourLearning page or attempting to address issues via a TechZone ticket will eventually get a response, but the above two methods are significantly preferred and will result in a much quicker resolution.

1. Use case

In this lab, your client is a large insurance company that is attempting to infuse AI into their business, while complying with best practices and regulations. Using the watsonx.governance platform, you will demonstrate how, with a single solution, they can govern, monitor, and document both traditional predictive models and new generative AI models.

2. Environment

To complete this lab, you will need an IBM ID and access to an IBM Cloud account with several services configured. There are two options for accessing an account.

Use a TechZone Account

This is the preferred method for completing the lab, and the one reflected in the lab instructions. Business partners and IBMers can use the [IBM watsonx.governance account reservation](#). Please note that these accounts have proved to be very popular, and can run into capacity limitations at times.

WHEN USING TECHZONE ACCOUNTS, DO NOT ATTEMPT TO PROVISION NEW INSTANCES OF SERVICES SUCH AS WATSON MACHINE LEARNING OR WATSONX.GOVERNANCE. Doing so will create lite versions of these, which will immediately run out of capacity.

Use your personal account

It is possible to use a personal Cloud account with the following services configured:

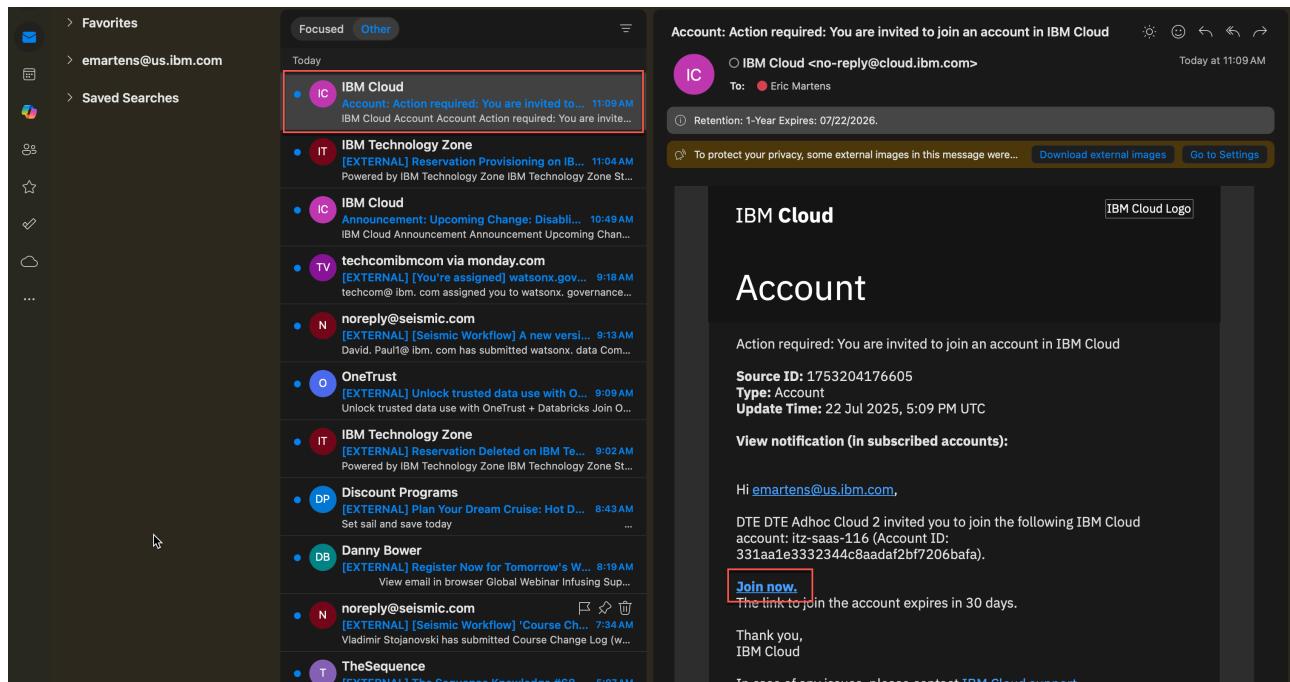
- [Watson Machine Learning](#)
- [watsonx.governance](#) (previously known as Watson OpenScale)
- [IBM Knowledge Catalog](#)
- [IBM Cloud Object Storage](#)

You may use the free "lite" versions of these services; however, performing the lab steps will quickly hit the monthly limits for compute usage, meaning you may struggle to complete the lab in a timely fashion. For this reason, we recommend using TechZone shared account.

Reserving a TechZone account

The following instructions in this section will assume that you are using a TechZone account environment; if you are using your own account, you can skip ahead to the [A note on platforms](#) section.

1. Click on the link for the [IBM watsonx.governance account reservation](#) and follow the instructions to make a reservation. The reservation will provide access to a Cloud account with the necessary services for a limited time. When your reservation is complete, you should receive an email notification inviting you to the account.



2. Click on the [Join now](#) link from the email. The invitation with the [Join now](#) link can also be found in your [IBM Cloud account notifications](#).
3. Follow the instructions to verify your identity and join the account. Be sure to note the account name, which you will need when switching between the TechZone account and your personal account.
4. If you have not already, sign into [IBM Cloud](#).
5. Use the [account switch dropdown](#) to verify that you are signed into the TechZone account for which you received the invitation.

The screenshot shows the IBM Cloud dashboard. At the top, there's a search bar and navigation links for Catalog, Manage, and a specific resource named '3115389 - itz-saas-116'. Below the search bar is a 'Dashboard' section with a 'For you' sidebar containing icons for Build, Track emissions with Carbon Calculator, Build with Watson, Use Watson Assistant, Use Speech to Text, and Create a project. There's also a 'Create resource' button.

You will need to launch the watsonx.governance service from the Cloud dashboard.

Configuring a TechZone account

TechZone accounts may require you to configure watsonx services and create an inventory.

1. Configure the monitoring service

First, you will ensure that the monitoring service (formerly known as OpenScale) is configured.

1. Click on the [Resource list](#) icon on the left. The [Resource list](#) screen opens.

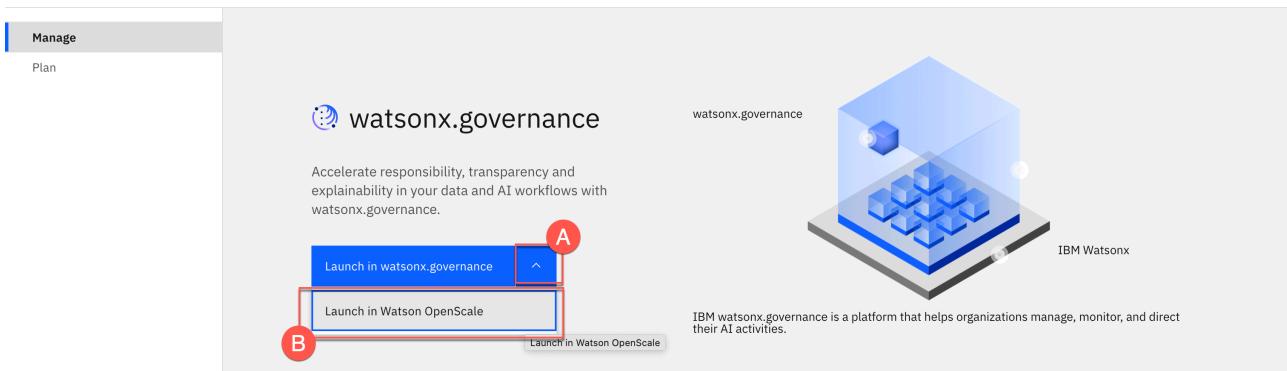
The screenshot shows the IBM Cloud dashboard again, but this time the 'Resource list' icon in the sidebar is highlighted with a red box. The rest of the interface is similar to the first screenshot.

2. Click on the [AI / Machine Learning](#) section (A) to open it. If you are using a TechZone account, take a moment to verify that the account contains [watsonx.governance](#) and [watsonx.ai Runtime](#) services. If either service is not available, you may need to delete and re-create your TechZone account reservation before you can proceed. Locate the [watsonx.governance](#) service (B) from the list and click on it. The [watsonx.governance](#) service screen will open.

The screenshot shows the 'Resource list' screen. The 'AI / Machine Learning' section (A) is expanded, showing three resources: 'gov-110000B3QC' (watsonx.governance), 'wml-110000B3QC' (watsonx.ai Runtime), and 'ws-110000B3QC' (watsonx.ai Studio). The 'gov-110000B3QC' row is highlighted with a red box (B). The 'watsonx.governance' service row has a red box around its name in the 'Product' column.

Name	Group	Location	Product	Status	Tags
gov-110000B3QC	itz-687fc47cbefc366aa98823c2	Dallas (us-south)	watsonx.governance	active	eid:687... +2
wml-110000B3QC	itz-687fc47cbefc366aa98823c2	Dallas (us-south)	watsonx.ai Runtime	active	eid:687... +2
ws-110000B3QC	itz-687fc47cbefc366aa98823c2	Dallas (us-south)	watsonx.ai Studio	active	eid:687... +2

3. Click on the [arrow icon](#) in the launch dropdown (A) to expand it, then click on the [Launch in Watson OpenScale](#) item. The Watson OpenScale interface opens in a new browser tab.



4. If the OpenScale service has not previously been configured, you will be asked if you would like to perform a manual setup or an auto setup, like the screenshot below. Manual setup is beyond the scope of this lab. Click on the **Auto setup** button to kick off the setup function, which will take roughly 30 minutes to run. Note that you can proceed with the remainder of the lab while this runs.

If OpenScale has already been configured, you will not see the setup options, and should be able to proceed with the lab.

The screenshot shows the IBM Watson OpenScale Insights dashboard. It features a central 3D isometric illustration of a model evaluation process. On the left, there are summary metrics: 'Deployments Monitored' (0), 'Quality Alerts' (--), and 'Fairness Alerts' (--). Below these are filters for 'Tags', 'Alert type', and 'Machine learning provider'. A search bar says 'Which deployment are you looking for?'. To the right, there are sections for 'Drift v2 Alerts', 'Drift Alerts', 'Global explanation Alerts', and 'Custom Alerts'. At the bottom, there are two buttons: 'Manual setup' (gray) and 'Auto setup' (blue, highlighted with a red box labeled 'B').

5. If the automatic setup succeeds, you can proceed to [the next step](#). However, if it fails, click on the [Manual setup](#) button.

This screenshot is identical to the previous one, showing the IBM Watson OpenScale Insights dashboard. The central 3D isometric illustration and surrounding UI elements are the same. The 'Auto setup' button is now grayed out and disabled, while the 'Manual setup' button is highlighted with a red box labeled 'B'.

6. Click on the [Edit icon](#) in the Database tile.

The screenshot shows the 'System setup' page in the IBM Watson OpenScale interface. On the left, a sidebar lists 'Required' and 'Optional' sections. Under 'Required', 'Database' is selected. The main panel displays the 'Database' configuration, which includes a 'Description' field containing the text: 'The Watson OpenScale database stores your model transactions and model evaluation results.' A blue box highlights the 'Database' configuration area.

7. Click on the **Database type** dropdown and select the **Free lite plan database** option. Click on the **Save** button to save your changes.

The screenshot shows the 'Database' configuration page. The 'Database type' dropdown is set to 'Free lite plan database', which is highlighted with a red box. The main panel contains a 'Description' field with the text: 'Select a database to store your model transactions and model evaluation results. A free database is available for Lite plan users to get started. Alternatively, you can use an existing PostgreSQL or Db2 database (lite Db2 plans not supported). You can also purchase a new database.' Below it is a note: 'Note: The free Lite plan database is not GDPR compliant. If your model processes personally identifiable information (PII), you must purchase a new database or use an existing database that does conform to GDPR rules. [Learn more](#)'.

Next, you will identify or create a model inventory.

2. Identify a model inventory

The watsonx.governance solution allows organizations to group and track their models based on use cases, or issues that models are attempting to solve. Each use case stores and organizes data and lifecycle information for candidate models in development, testing, and production phases.

Note: In a real world use case, an organization would identify a business need for an AI model, and create a use case to track that effort. These use cases are fully integrated into IBM OpenPages, a full risk management solution, as part of a broader AI governance process. The [watsonx.governance Level 4 PoX lab](#) covers these capabilities in greater depth.

These use cases are collected as assets in a model inventory. Access to a model inventory is required to complete this lab. Follow the steps below to determine if you already have access to an inventory.

1. Return to the browser tab showing the watsonx.governance resource in IBM Cloud, or navigate back to it from the [IBM Cloud Resources](#) page. Ensure that you have switched to the correct account using the account switcher.
2. Click on the [Launch in watsonx.governance](#) button. The watsonx home screen will open.

watsonx.governance

Accelerate responsibility, transparency and explainability in your data and AI workflows with watsonx.governance.

Launch in watsonx.governance

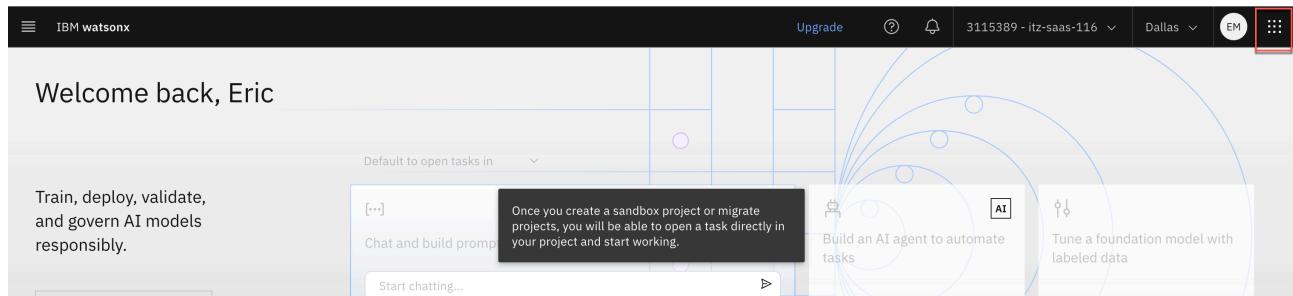
Helpful links

IBM watsonx.governance is a platform that helps organizations manage, monitor, and direct their AI activities.

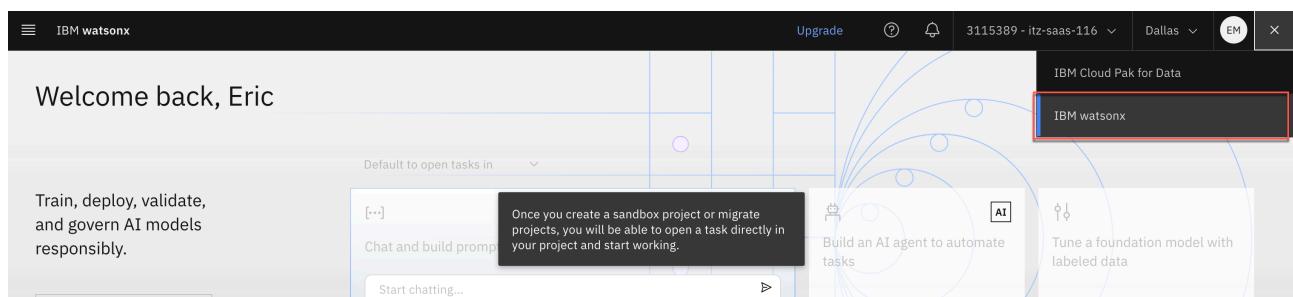
Note: The first time you launch the service, you may need to accept Data Use Policy terms. Close any windows offering to let you take a tour, or dive deeper into the product.

A note on platforms

All the instructions and screenshots in this lab will use the new **watsonx** platform, screens, and tooling. The traditional Cloud Pak for Data functionality will be combined with the watsonx functionality over the course of 2024. If your screen looks different than the ones shown here, you may need to switch to the new context by clicking the **Switch platforms** button at the top right of your screen:

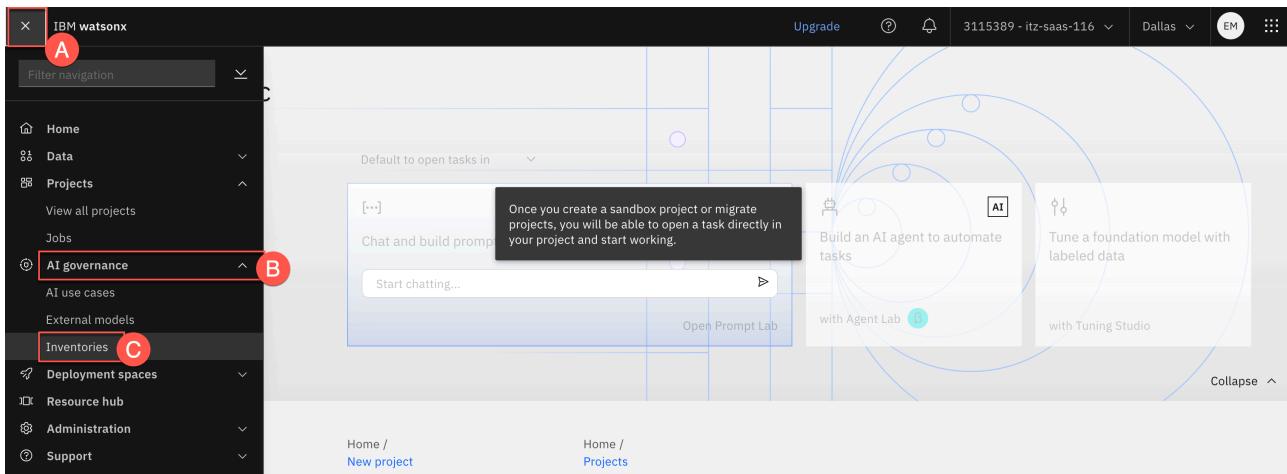


Click on the **IBM watsonx** option from the dropdown menu to select the watsonx platform:



Note that the URL in your browser will reflect the platform you have chosen. If the URL ends in `context=watsonx`, you are using the watsonx platform. If it ends in `context=cpdaas`, you are using the Cloud Pak for Data platform.

3. Click on the **navigation menu** (A) in the upper left to expand it. Locate the **AI governance** section (B) of the menu, expanding it if necessary, and click on **Inventories** (C). The Inventories screen opens.



IF YOU DO NOT HAVE AN AI GOVERNANCE MENU OPTION AVAILABLE, IT IS HIGHLY LIKELY THAT YOU HAVE LAUNCHED THE SERVICE IN THE WRONG ACCOUNT. It is also possible that your login has expired. You may need to log out and log in again, or potentially clear your browser cache.

If you see a warning message about [AI use cases deprecated as part of COS lite plan](#), you can safely ignore the message.

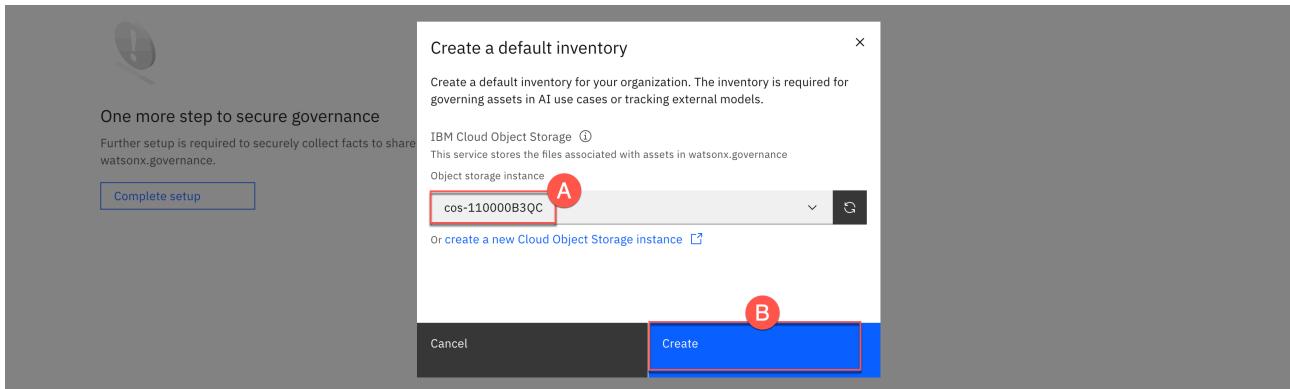
4. Verify that you have an inventory available to you in the list.

Name	Date created	Creator	Your role
itz-inventory-watsonx-036-fra	3 weeks ago	Ben Foulkes	Editor

If you **DO NOT** have an inventory, you will likely need to create one.

5. Click on the [Complete setup](#) button to set up the watsonx.governance service. The [Create a default inventory](#) popup opens.

6. Click on the [Object storage instance](#) dropdown (A) and select the storage service associated with your account. Click on the [Create](#) button (B) to create a default inventory for the use cases.



If you now have access to a model inventory, you can skip to the appropriate section:

- To learn how to apply AI governance principles to predictive models, go to the [Governing predictive models](#) section.
- To learn how to apply AI governance principles to generative AI models, go to the [Governing generative models](#) section.

Governing predictive models

The release of ChatGPT ensured that generative AI and large language models (LLMs) dominated the recent AI conversation. However, most organizations are still working to integrate traditional predictive models into their businesses. These efforts are complicated by ever-increasing regulations and public scrutiny. While predictive machine learning models may not receive the same attention as their generative counterparts, there is still a tremendous opportunity to help clients with these use cases.

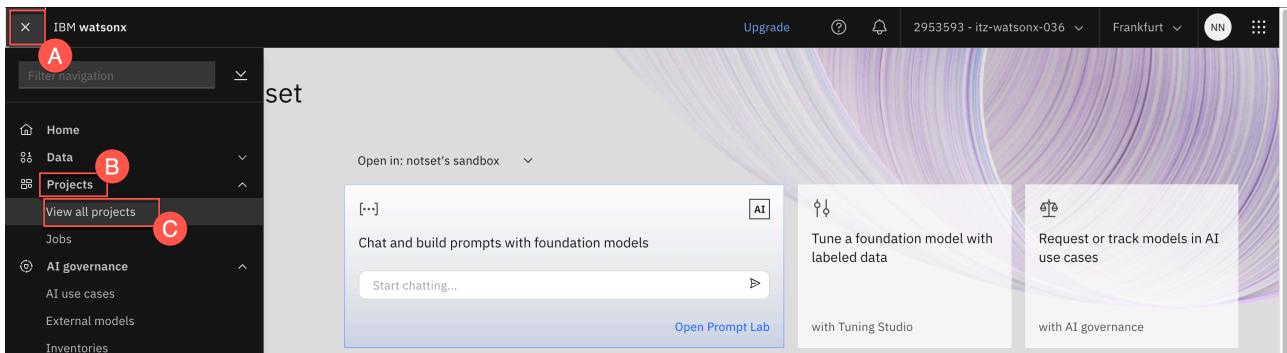
In this section of the lab, you will work with a model to predict the risk associated with auto insurance policies based on various facts about the policyholder. You will track that model through its lifecycle, from the development phase through the testing and deployment phases. You will configure the model for monitoring, and see how the model evaluations are performed and recorded using the watsonx.governance platform.

1. Set up a watsonx project

In this section, you will create an IBM watsonx project that will contain all the assets used to deploy and work with the predictive model. Watsonx projects provide a central location for data scientists, data engineers, subject-matter experts, and other stakeholders to collaborate on data science projects. This lab uses only a few of its capabilities, which include support for industry standard development environments such as Jupyter notebooks and R Studio, easy connections to hundreds of different types of data sources for drag-and-drop insertion of data into code, visual model building and data manipulation with IBM's SPSS Modeler, synthetic tabular data generation, and more.

Create the project

1. Right-click link for the [project file](#) and choose the appropriate menu option for your browser to download it to your machine. Do **NOT** unzip the file.
2. Navigate to the IBM watsonx projects screen. As always, ensure that you are using the correct account and region. You can reach the projects screen by click on the [Navigation menu](#) in the upper left (A), clicking on the [Projects](#) item (B) to expand it, then clicking on the [View all projects](#) menu item (C).

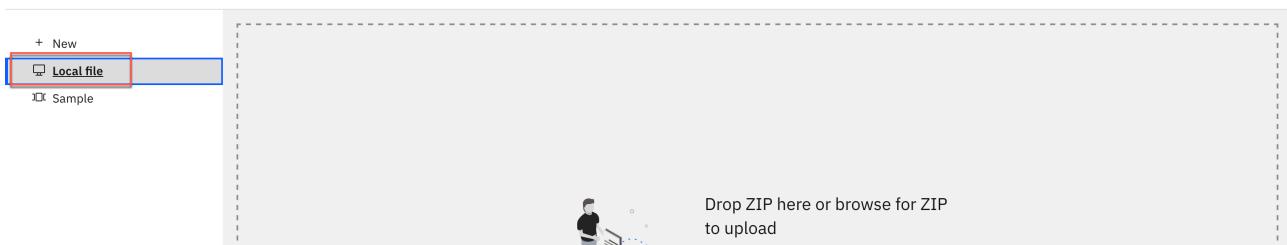


3. Click the blue **New project** button on the right.

4. Click the **Local file** option on the left.

Create a project

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



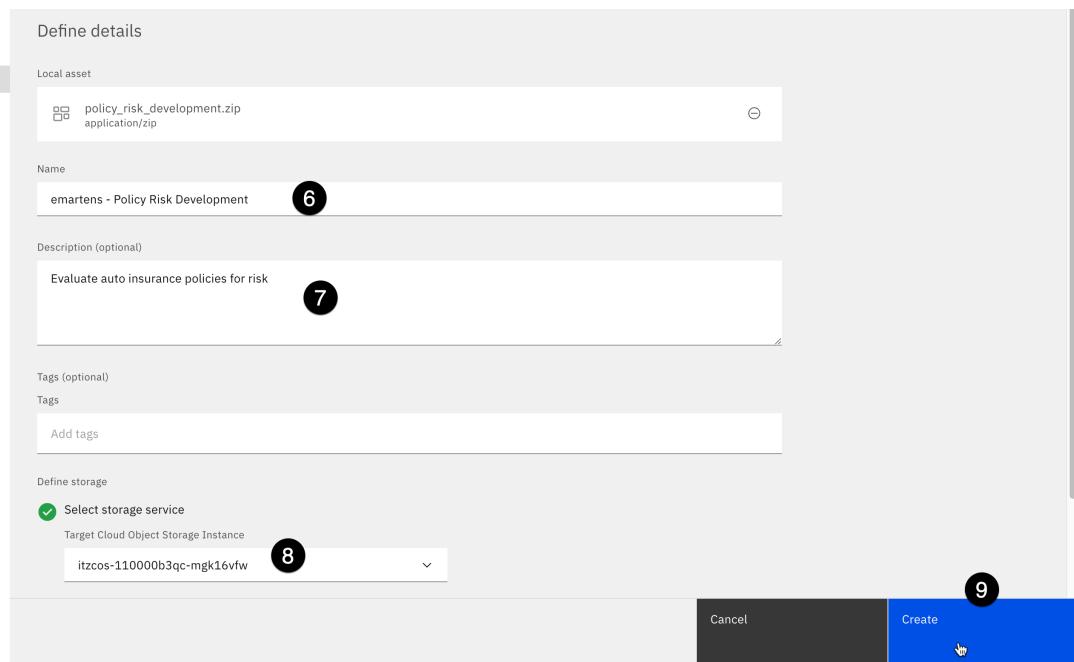
5. Click the **Browse** button in the middle of the screen, and browse to the zipped *policy_risk_development.zip* file you downloaded in step one.

6. Give your project a **Name** with identifying information such as **your email address and Policy Risk Development**.

7. Give your project a **description**.

8. Click on the **Select storage service** dropdown and select the storage service from your reserved environment. If there are multiple instances, choose the one that most resembles the service in the screenshot. If there are no services here, **DO NOT ADD OR CREATE A NEW OBJECT STORAGE INSTANCE**. Instead, try refreshing the page, clearing your browser cache, or logging out and logging back into watsonx until the associated object storage instance appears.

9. Click on the **Create** button to create the project from a file.



Verify and configure the project

When importing a project from a file, it's critical to ensure that all the resources import successfully, since the project tool will report that creation was successful even if one or more resources fails to import.

1. Click the [View import summary](#) button in the Project History tile and ensure that nothing is listed in the [Incomplete](#) or [Failed](#) categories on the left of the screen. Note that as capabilities change and this lab evolves, your screen may show more assets being imported. The important thing is that all included assets are imported, and that there are no Incomplete or Failed entries.

If an asset failed to import, you will need to return to the [IBM watsonx projects screen](#). Locate the project from the list, check the box to the left of it, and then click the [Delete](#) button from the blue menu bar that appears above the table. Then repeat the section above to re-create the project. The vast majority of project import failures can be solved by deleting and re-importing the project.

2. Once the project has successfully imported, click the [Close](#) button.

emartens - Policy Risk Development

Project Import Summary

The project imported successfully.

Import File	Import Date	Imported By
policy_risk_development.zip	Sep 17, 2024	Eric Martens

Q: What assets are you looking for? All asset types ▾

Imported 4	Name	Size	Type
policy_risk_eval.csv	15.93 KB	Data Asset	
policy_risk_train.csv	7.08 MB	Data Asset	
Policy Risk Prediction	---	Pipeline	
Policy Risk Prediction - P4 Ridge - Model	8.2 KB	Model	

Incomplete 0

Failed 0

Close 2

3. Click on the **Manage** tab.
4. Click the **Services & integrations** item from the menu on the left.
5. Click the blue **Associate service** button on the right. A list of services will appear in a table.

IBM watsonx

Projects / emartens - Auto policy risk

Overview Assets Jobs Manage 3

Services & integrations

IBM services Third-party integrations

Find services

Associate service + 5

4

6. Locate the appropriate **watsonx.ai Runtime** (formerly Watson Machine Learning) service for the account in the table. The **Type** column should say **watsonx.ai Runtime** or similar. If you have multiple options, choose the one that uses the **Essentials** or **Standard** plan. Check the box to the left of the service.

Note: If no service is listed, **DO NOT** create a new one. Please see the **Troubleshooting** section for potential fixes for this issue.

7. Click on the blue **Associate** button.

The project is now configured and ready to use.

2. Define a model use case

In this step, you will create a use case for the predictive risk model in the model inventory you identified (or created) in the previous step. In this case, an insurance company would like to assess the risk associated with auto insurance policies. Their data science team has gathered data on various accident "hotspots" in metro Chicago, where traffic accidents happen more frequently. The data scientists have found that the closer a policyholder lives to a hotspot, the more likely they are to be involved in an accident and file a claim. They have suggested incorporating this data with other known risk factors such as driver age, gender, and vehicle type to build an AI risk prediction model.

As the first part of the process for approving the model, the stakeholders in the company would create a use case for the policy evaluation model to organize the different versions, data, and environments used to build and deploy the candidate models.

Define the use case details

1. Verify that you are signed into the IBM watsonx service using the correct account and region.
2. Click on the **Navigation menu** in the upper left (A) to expand it. Locate the **AI governance** section of the menu (B), expanding it if necessary, and click on the **AI use cases** menu item (C).

4. Click the New AI use case button to open the New AI use case window.

The screenshot shows the 'AI use case' section of the IBM Watson interface. At the top right, there is a blue button labeled 'New AI use case'. The main area displays a table with columns: Name, Status, Owner, Inventory, Tags, Risk level, and Alerts in. One row is visible, showing 'Claim summarization' as the name, 'Development in progress' as the status, 'Eric Martens' as the owner, 'Default Inventory' as the inventory, 'Medium' as the risk level, and 'None' as the alerts.

5. Give your use case a name. If you are using a shared account, use some identifying information such as [your](#) email address to mark it as belonging to you.

6. Give your use case a description for the business issue it is attempting to solve.

7. Use the [Risk level](#) dropdown to set the associated level of risk. In a real-world example, this would be performed by the risk management officer of an organization. Environments integrated with IBM OpenPages will offer more risk management functionality.

8. Use the [Inventory](#) dropdown to select the model inventory you identified in a previous step.

This screenshot shows the 'New AI use case' dialog. On the left, there's a sidebar with tabs for 'Narrative' and 'AI'. The main area has sections for 'General information' (Name: 'emartens - Auto policy risk' (5), Description: 'Evaluate the risk for auto insurance policies' (6)), 'Risk level' (Medium (7)), and 'Inventory/Catalog' (Required: 'emartens - Auto insurance (itz-watsonx-6)' (8)). On the right, there are sections for 'Owner' (Eric Martens), 'Status' (Draft), and 'Tags' (Search tags). A note at the bottom says 'Add tags to this AI use case.'

9. Click on the [Status](#) dropdown on the right side of the screen, and note the different values available.

This screenshot shows the 'New AI use case' dialog with the 'Status' dropdown open. The dropdown menu lists several options: 'Development in progress' (1, highlighted), 'Awaiting development', 'Developed', 'Ready for AI asset validation', and 'Validation complete'. The rest of the dialog and sidebar are similar to the previous screenshot.

Watsonx.governance allows for organizations to implement formal approval procedures around their model lifecycles. In this case, the auto policy team has requested development of a new model to take into account information gathered by their data scientists. They create a use case for the business problem to request further exploration and development (the *Ready for use case approval* status), and can update the status of the use case as it moves through the process, from initial approval (the *Use case approved* status) to the assigning of AI engineering resources (the *Awaiting development* status). For this lab, you will move directly to an approved use case.

10. Use the dropdown to set the [Status](#) to [Development in progress](#).

11. Click on the [Create](#) button. After a moment, the use case will be created.

12. Take a moment to review the use case screen, and note the [Access](#) tab, which allows sharing of the use case with other stakeholders to allow collaboration on the model lifecycle.

You have successfully created a use case for an AI model to address a business need. In the next step, you will associate workspaces with the use case.

3. Associate a project with the use case

In this step, you will take on the role of an AI engineer, and prepare for developing a model. In watsonx, model development is done in a project. Projects contain models, data assets, metadata, and access controls necessary for AI governance. They offer full access management, and allow data scientists, data engineers, and others to collaborate on AI and machine learning projects. In the context of an AI use case, they can be described as [workspaces](#). Associating workspaces with different phases of the model lifecycle in the use case allows you to organize and manage governance materials in a logical manner, linking the projects created to the relevant AI use case.

1. From the [Overview](#) tab of the use case details screen, scroll down to the [Associated workspaces](#) section and click on the [Associate workspaces](#) button in the [Development](#) tile. The [Associate workspaces](#) window opens.

The screenshot shows the 'AI use cases / emartens - Auto policy risk' screen. The 'Overview' tab is selected. On the left, there's a 'Associated workspaces' section with a note: 'Associate your AI use case with the workspaces in order to organise them under the same business problem.' Below it are two boxes: 'Develop' (with a note about ML development) and 'Validate' (with a note about pre-production deployment). A blue button labeled 'Associate workspace +' is highlighted with a circled '1'. To the right is a circular diagram showing 'Workspaces for Develop' connected to an 'AI use case'. On the far right, the 'About this AI use case' panel displays details: Name (emartens - Auto policy risk), Description (Evaluate the risk for auto insurance policies), Owner (Eric Martens, emartens@us.ibm.com), Status (Development in progress, Sep 17, 2024), Risk level (Medium), and Inventory/Catalog.

Workspaces associated with AI use cases can contain either projects, as discussed above, or spaces, which will be discussed in the next section. If you wish, you can use the [Read more](#) links on this screen to find out about the rules and reasoning behind associating workspaces with phases of the model lifecycle. Note that multiple workspaces can be associated with a single phase of the lifecycle, but a workspace can only be associated with one lifecycle phase. For example, the data science team may be working on several different candidate models for a use case, each with their own assets, collaborators, and datasets. All of these projects would be associated with the *Development* phase of the single use case. However, each development project would only be associated with the *Development* phase, and could not be associated with the *Validation* or *Operation* phases.

2. Scroll down to the [Projects](#) section of the window. Any projects that you have access to will be listed here. Locate the [Policy Risk Development](#) project you created in the previous step, and check the box

to the left of it.

3. Click on the [Save](#) button to associate the project with the use case.

The screenshot shows a table with columns: Name, Created, Your role, Tracked AI assets, and Associates. The first row has a checked checkbox in the Name column, and the 'emartens - Policy Risk Development' project is selected. The 'Associates' column shows a value of 0. At the bottom, there are buttons for 'Cancel' and 'Save'. The 'Save' button is highlighted with a blue background and a circled number 3.

The [Associate workspaces](#) window will be updated to show the new project associated with the AI use case in the [Development](#) tile. You have now created a project for the development of the risk assessment model. In the next section, you will create a space so that the developed models can be tested.

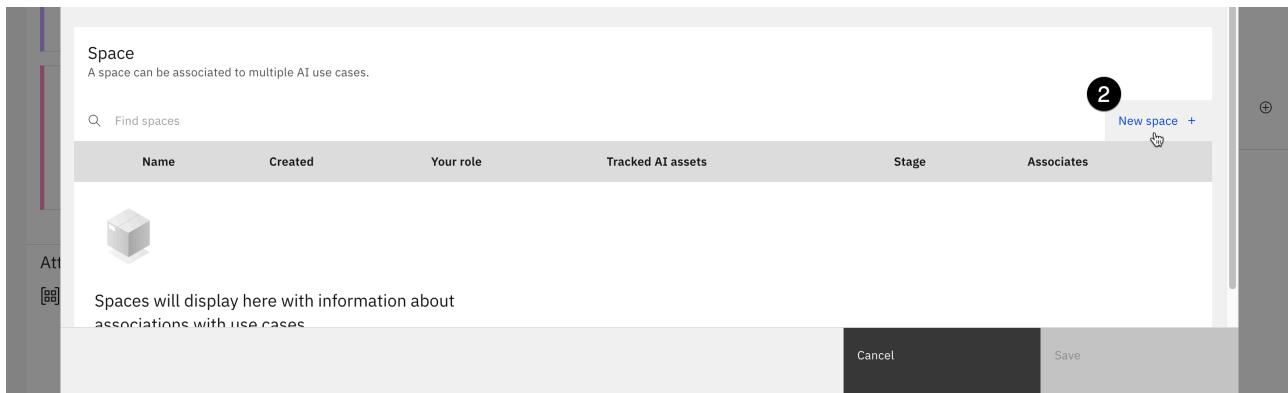
4. Associate a space with the use case

[Spaces](#) contain deployable assets, deployments, deployment jobs, associated input and output data, and runtime environments. You can use spaces to deploy various assets and manage your deployments and control access to the models and data stored there. Later in the lab, you will promote your developed model to a space for testing. To ensure that the test results for the model are reflected in the AI use case, you will need to associate those spaces with the AI use case.

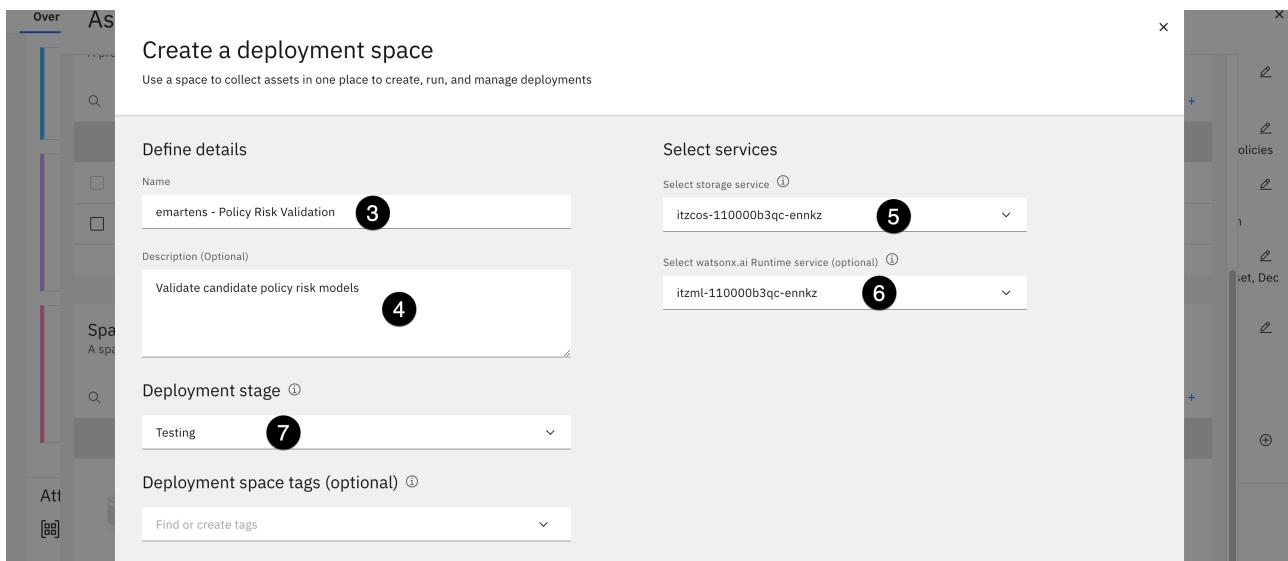
1. From the [Associated workspaces](#) section of the use case screen, locate the [Validation](#) tile and click on the [Associate workspaces](#) button.

The screenshot shows the 'Associated workspaces' section of the use case screen. The 'Validate' tile is active, showing an 'Associate workspace +' button with a circled number 1. To the right is a circular diagram illustrating the relationship between the AI use case and three workspaces: Workspaces for Develop, Workspaces for Validate, and Workspaces for Operate. On the far right is the 'About this AI use case' panel, which contains details such as Name, Description, Owner, Status, Risk level, Inventory/Catalog, Tags, and creation/modification history.

2. Scroll down to the [Space](#) section of the [Associate workspaces](#) window and click on the [New space](#) button. The [Create a deployment space](#) window opens.



3. Give your space a **Name** with identifying information such as [your email address](#) and [Policy Risk Validation](#).
4. Give your space a [Description](#).
5. Click on the [Select storage service](#) dropdown and select the service associated with your reservation. Again, if you have multiple options, select the one that most resembles the screenshot.
6. Click on the [Select watsonx.ai Runtime service](#) dropdown and select the environment's watsonx.ai Runtime service from the list. Note that if the object storage instance or the machine learning service instance does not appear, [DO NOT ADD OR CREATE NEW INSTANCES](#). Instead, try refreshing the page, clearing your browser cache, or logging out and logging back into watsonx until the associated instances appear. If you have multiple instances, select the one you used when creating your project.
7. Click on the [Deployment stage](#) dropdown and select [Testing](#) to specify that this space will be used for validation testing.



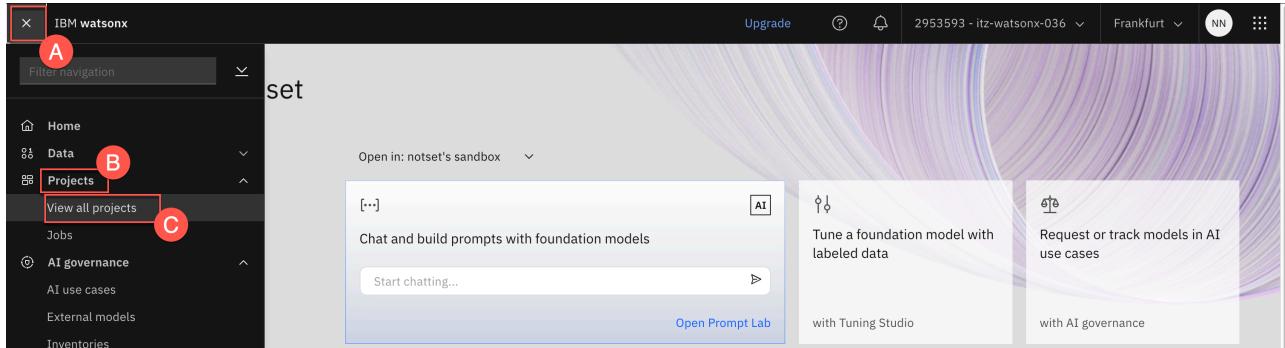
8. Click on the [Create](#) button in the lower right to create the space. Space creation can take up to two minutes to complete. When it finishes, a notification window will appear telling you that the space is ready. Click the [X](#) button to close the notification window and return to the [Associate workspace](#) window, which should now show the newly created space.
9. Click on the [Save](#) button in the lower right to save the workspace association and return to the AI use case overview screen, which will now show the space listed in the [Validate](#) tile of the [Associated workspaces](#) section.

5. Track the model

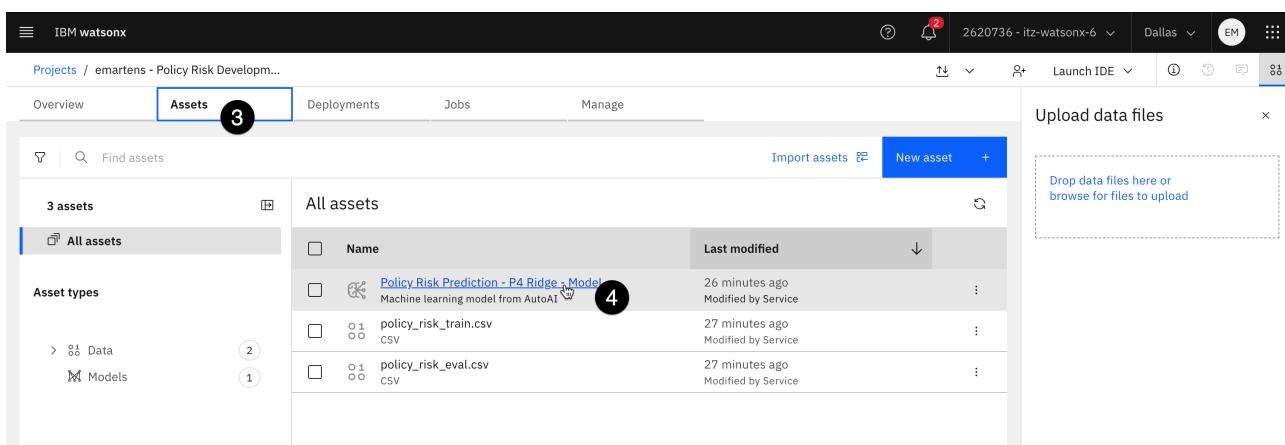
This lab will use a model built with AutoAI, IBM's rapid model prototyping service that can quickly generate predictive machine learning models from tabular data and save the output as either a Jupyter notebook, or as a ready-to-deploy model.

Configure model tracking

1. Return to the IBM watsonx projects screen. As always, ensure that you are using the correct account and region. You can reach the projects screen by click on the **hamburger menu** in the upper left, clicking on the **Projects** item to expand it, then clicking on the **View all projects** menu item.



2. From the list of projects, click on the **Policy Risk Development** project you created in a previous step to open it.
3. Click on the **Assets** tab of the project. Note that the *policy_risk_train.csv* training data file has been provided if you would like to run your own AutoAI experiment, or use a different method to create the model.
4. From the list of assets, locate and click on the **Policy Risk Prediction - P4 Ridge - Model** entry to open the model information screen.



5. Take a moment to review the information presented. The metadata includes when the model was created, the identity of the creator, the prediction type, algorithm used, and information on the training dataset.
6. Scroll down to the **Training evaluation** section and note that the initial quality metrics generated by AutoAI during model creation have been captured here. Finally, note that the model's input schema is included.

As regulations around the use of AI models increase, the ability to automatically track and easily retrieve model metadata without any manual effort from data scientists has become critical. Watsonx provides this capability, improving transparency and helping speed time-to-value for AI models.

7. Scroll back up to the **Governance** section at the top of the model information screen and click the **Track in AI use case** button. The **Track asset** screen opens.

The screenshot shows the IBM WatsonX interface. On the left, there's a sidebar with navigation items: Governance, Model, Development (with sub-items emartens - Policy Risk Development, Training evaluation, Additional details, Attachments), and a search bar. The main content area has a title 'Governance' with a small icon. Below it, a message says 'This asset is not tracked.' and 'To track an asset, add it to an AI use case. Tracking captures details about the asset for governance purposes.' At the bottom of this section is a blue button labeled 'Track in AI use case'. To the right, there's a panel titled 'About this asset' with fields for Name (Policy Risk Prediction - P4 Ridge - Model), Description (Predict auto policy risk), Asset Details (Type: wml-hybrid_0.1, Model ID: 331bd3af-99cd-4d...), and Tags (Add tags to make assets easier to find). There are also 'Export report' and 'Edit' buttons.

At this point, if you receive a message informing you that your project has not been associated with a use case, you will need to return to the use case and associate the project as a workspace, as described in the [Define a mode use case](#) section.

8. When asked to define an approach, leave **Default** approach selected and click the **Next** button.

9. When asked to assign a model version, leave **Experimental** selected. Note that you can manually assign a version number here, or choose a more production-ready version number depending on the state of the model. Click the **Next** button. The **Review** panel opens.

10. Verify that all details are correct, and click the **Track asset** button to begin tracking the model.

The screenshot shows a 'Review' panel. On the left, there's a sidebar with three options: 'Define approach' (radio button selected), 'Assign version' (radio button), and 'Review' (radio button). The main content area has a title 'emartens - Auto policy risk'. Below it, there are two columns: 'Approach' (set to 'Default approach') and 'Version' (set to '0.0.1'). A note below the approach says 'A default approach for tracking your AI assets.' and shows a long GUID: '00000000-0000-0000-0000-000000000000'. At the bottom, there are 'Cancel', 'Back', and a large blue 'Track asset' button. The 'Track asset' button is highlighted with a blue box and the number 10.

View the tracked model

Once the model tracking has been enabled, you will be returned to the model information screen. You can now view the model information in the use case.

1. Click the **View details** arrow icon button. A new tab will open in your browser showing the model use case in your model inventory.

2. Click on the [Lifecycle](#) tab of the use case.

3. Scroll down to the map of models contained in the [Default approach](#) section. Note that there are three lifecycle sections listed (*Development*, *Validation*, and *Operation*). The tracked model has not yet been promoted to a deployment space, so it is listed in the *Development* section of the lifecycle.

If you click on the model name from this screen, you will see the full model information that is also available from clicking on it from the project. The [Lifecycle](#) tab of the AI use case is useful for stakeholders and business users to get a quick overview of all the models attempting to address a particular issue, with the ability to drill down for more information as necessary, all without needing access to the project where the developers and data engineers are working.

You have successfully enabled model tracking, and can now observe the changes as the model goes through the lifecycle.

6. Deploy the model

Promote the model to the deployment space

Next, you will promote the model to a deployment space and deploy it. Watsonx uses deployment spaces to organize models and model-related assets for validation and production access. Deploying the model allows it to be accessed via REST API for further testing.

1. Return to the browser tab showing the model in the project. You can do this by returning to the previous browser tab, or by navigating back to your project list, then clicking on your project.
2. From the list of assets, click on the [three vertical dots](#) to the right of the model to open the context menu.
3. Click on the [Promote to space](#) item from the menu to promote the model to a deployment space. The [Promote to space](#) screen opens.

4. Click on the **Target space** dropdown, then click on the space you created in a previous step.
5. Check the box next to [Go to the model in the space after promoting it](#).
6. Click **Promote**. The deployment space screen will load, with the entry for the model open.

Create a deployment

The model has been promoted to the space. Deployment spaces can contain different types of assets, including the models themselves, as well as data used in batch processing jobs. Spaces are also fully governed, allowing administrators to provide different levels of access for different stakeholders.

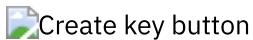
Models in a space can have online and batch deployments created that allow access to the models.

1. Click the **New deployment** button. The [Create a deployment](#) screen will open.

2. A warning will likely appear informing you that you do not have a user API key. Click the [Create](#) button in this warning. A new browser tab will open to the [User API key](#) screen.



3. Click the [Create a key](#) button. A new API key will be created for your user. When it is finished, you will see an API key in the table of active keys.



4. Return to the browser tab with the [Create a deployment](#) screen open, and click the [Reload](#) link in the [No user API key](#) warning banner. Your new API key will be verified, and then the warning banner will disappear.



5. Ensure that the [Online](#) tile is selected. This deployment type will allow for REST API access.

6. Give your deployed model a name with personally identifiable information, such as [your](#) email address.

7. Click the [Create](#) button. Deploying the model will take roughly a minute.

The screenshot shows the 'Create deployment' form. The 'Deployment type' section has two options: 'Online' (selected) and 'Batch'. The 'Online' option is described as 'Run the model on data in real-time, as data is received by a web service.' The 'Batch' option is described as 'Run the model against data as a batch process.' The 'Name' field is filled with 'emartens - Policy risk testing'. The 'Serving name' field is empty. The 'Description' field contains 'Deployment description'. The 'Tags' field is empty. At the bottom right, there are 'Cancel' and 'Create' buttons, with the 'Create' button being highlighted with a circled number 7.

View the changes in the lifecycle

The model's deployment to the space associated with the [Validation](#) phase of the lifecycle in the AI use case will cause updates to its status in the use case, as it has now progressed in the lifecycle.

1. When the deployment is finished, the *Status* in the displayed table will change to [Deployed](#). Once this has happened, click on the name of the deployment. The deployment details screen opens.

The screenshot shows the IBM Watsonx interface with the Deployments tab selected. A deployment named "emartens - Policy risk testing" is listed as "Online" and "Deployed". A red circle with the number 1 is on the deployment row. To the right, there is an "About this asset" panel with details like Name (Policy Risk Prediction - P4 Ridge - Model), Description (Predict auto policy risk), and Asset Details (Type: wml-hybrid_0.1, Model ID: 479db39c-56c8-4a...). A blue button at the top right says "New deployment".

2. Note the details provided on the [API Reference](#) tab include direct URLs to the model, as well as code snippets in a variety of programming languages that allow application developers to seamlessly include the model in their apps. The [Test](#) tab allows for direct testing of the model using either uploaded CSV files or JSON input.

3. Click on the [AI Factsheet](#) tab. The AI Factsheet for the model will load.

The screenshot shows the AI Factsheet tab selected. It displays deployment details for "emartens - Policy risk testing" (Deployed, Online). The "API reference" tab is selected, showing Direct link and Public endpoint URLs. A red circle with the number 3 is on the tab bar. To the right, there is an "About this deployment" panel with details like Name (emartens - Policy risk testing), Description (No description provided), and Deployment Details (Deployment ID: 69d46659-752a-4292-b077-1b19b86b8d72, Serving name: No serving name, Software specification: hybrid_0.1, Hybrid pipeline software specifications: autoai-kb_rt23.1-py3.10, Copies: 1).

Note that the model metadata has carried over from the tracking established in the project.

4. Scroll to the bottom of the factsheet. In the [Interested in more details?](#) tile, click on the arrow icon to open the model factsheet.

The screenshot shows the "Interested in more details?" tile at the bottom of the factsheet. It contains a message: "This information is part of AI factsheet. Click here to view the more details." A red circle with the number 4 is on the "More details" button. To the right, there is a "Copies" section with 1 copy, a "Serving name" section (No serving name), and a "Description" section (No description provided).

Note that the factsheet now contains sections for [Development](#) as well as [Validate](#). The [Development](#) section contains information from the project, such as the training data source and the initial evaluation from the AutoAI training. The [Validation](#) section contains information from the deployment space, including a link to the space and the ID of the deployment. As the model has not been evaluated in this state yet, no evaluation metrics are available.

The next step will be to validate that the model is making fair, accurate decisions.

7. Evaluate the model

In this section, you will evaluate the model for quality and fairness.

Enable the monitoring service

1. From the model's factsheet, click on the [Deployments](#) tab in the upper left.

Deployments / emartens - Policy Risk Validation / Policy Risk Prediction - P4 Ridge - Model

1 Deployments 2 AI Factsheet

Governance Model Development emartens - Policy ...

Governance

About this asset

Name: Policy Risk Prediction - P4 Ridge - Model
Description: Predict auto policy risk

2. Locate the deployment from the list and click on the name to open it.

Deployments / emartens - Policy Risk Validation / Policy Risk Prediction - P4 Ridge - Model

Deployments 2 AI Factsheet

Search

Name	Type	Status	Tags	Last modified
emartens - Policy risk testing	Online	Deployed	Add tags +	8 minutes ago Eric Martens (You)

New deployment

About this asset

Name: Policy Risk Prediction - P4 Ridge - Model
Description: Predict auto policy risk

Asset Details

Type: wml-hybrid_0.1
Model ID: 479db39c-56c8-4a...
Software specification:

3. Click on the **Evaluations** tab.

4. Click on the **Configure OpenScale evaluation settings** button. The **Prepare for evaluation** window opens.

emartens - Policy risk testing Deployed Online

Evaluations 3 Transactions AI Factsheet

Evaluation results
View Watson OpenScale evaluation results to gain insights about your model performance

Start configuring model evaluations

To view evaluation results, configure evaluations to monitor the deployment for fairness or performance drift.

Configure OpenScale evaluation settings

5. Note that the **Data type** and **Algorithm type** fields should automatically be set to their correct values. Click on the **View summary** button to proceed.

Prepare for evaluation

Provide deployment details to determine the appropriate configuration settings.

Details

Specify the data type and algorithm type for the model and provide the endpoint location if available.

Import settings

Data type
Numeric/categorical

Algorithm type
Regression

Model endpoint
<https://private.us-south.ml.cloud.ibm.com/ml/v4/deployments/19ea1f13-2596-4068-a744-eb355b0be395/preview>

Cancel Back View summary 5

6. The model summary screen contains information about the deployed model, including the input data type, label column, and more. This metadata has been automatically imported from the model's factsheet. Click on the **Finish** button to finish the setup and save the monitoring information. Note that the save operation may take up to two minutes to complete.

Connect to the training data

Next, you will configure the individual monitors for the model. Note that each deployed model can have its own custom metrics and alert thresholds configured, allowing administrators, compliance officers, and risk management professionals to ensure that the models meet all relevant regulations and internal requirements.

1. Right-click on [this link to the training data file](#) and download it to your machine. Note that this file is also available in the project you are using for this lab. Ensure that your browser does not change the extension to `.txt` when saving to your machine.
2. Click the **Edit** icon in the **Training data** tile.

Configure OpenScale evaluation settings

emartens - Policy risk testing

Model info

- Model details** (selected)
- Model governance
- Endpoints
- Explainability
- General settings
- SHAP
- LIME (enhanced)
- Evaluations
- Fairness

Model details

Description

Provide information about the training data and deployed model output to prepare Watsonx for monitoring and providing explanations for model transactions.

Reconfigure model

Training data

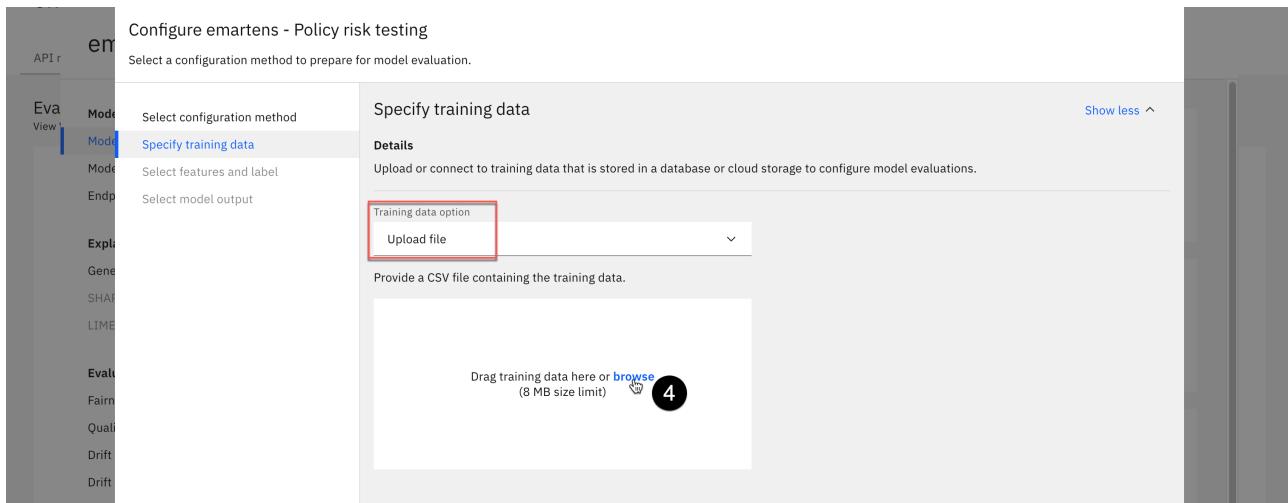
Storage type
Do not analyze

Training data label

Label column
RISK

2

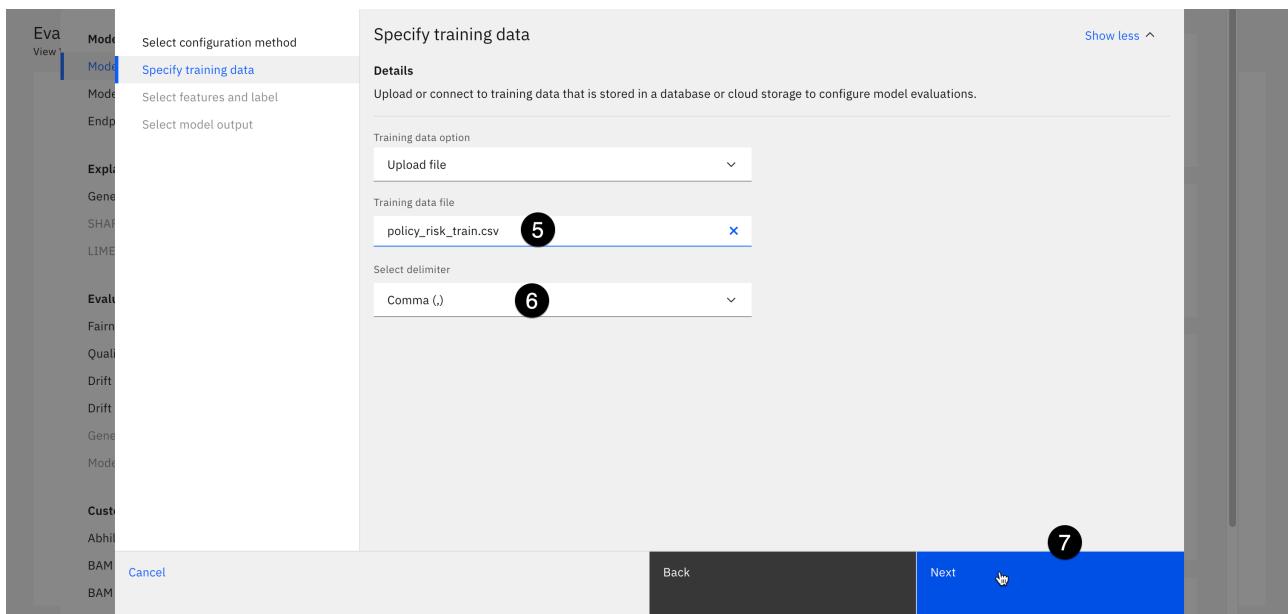
3. Leave the **Use manual setup** option selected for **Configuration method**, and click **Next**. The **Specify training data** screen opens.
4. Leave **Upload file** selected in the **Training data option** dropdown, and click on the **browse** link.



5. Use your machine's file manager to locate the *policy_risk_train.csv* training data file you downloaded in step 1.

6. Click on the **Select delimiter** dropdown and select the **Comma** option.

7. Click on the **Next** button. The monitoring service will load the training data file, and identify which columns represent features, and which represent output.



8. Scroll down in the table and verify that all of the columns are selected as **Features** EXCEPT for **RISK**, which should be selected as **Label/Target**.

9. Click on the **Next** button.

Select the feature columns and label column

Details

From the training data, select the input features. Indicate if numeric feature values represent categories by checking the Categorical checkbox. Next, select the label column. The label represents the correct prediction (ground-truth) for each record. A column can be marked as a feature or label but not both.

Select the feature columns

Column	Type	Categorical	Selected
INITIAL_ODOMETER	Number	<input checked="" type="checkbox"/>	<input type="checkbox"/>
LOW_MILEAGE_USE	Number	<input checked="" type="checkbox"/>	<input type="checkbox"/>
NB CLAIMS	Number	<input checked="" type="checkbox"/>	<input type="checkbox"/>
CLAIM_TOTAL	Number	<input checked="" type="checkbox"/>	<input type="checkbox"/>
NB_PRIM CLAIMS	Number	<input checked="" type="checkbox"/>	<input type="checkbox"/>
PRIM_CLAIM_TOTAL	Number	<input checked="" type="checkbox"/>	<input type="checkbox"/>
MINORITY	Text	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
RISK	Text	<input type="checkbox"/>	<input checked="" type="checkbox"/>

Items per page: 25 | 1 - 21 of 21 items

Selected features: 20

Back | Next

10. Verify that the [prediction](#) field is accurately identified as the model's [Prediction](#), then click on the [View summary](#) button to proceed.

11. Click [Finish](#) to save the training data setup. The service will upload the training data file, which make take a few minutes to complete.

Configure the fairness monitor

Next, you will configure the fairness monitor and set alert thresholds. Note that, for the sake of simplicity, this model includes the [MINORITY](#) field as a feature in the model. In a real-world example, this would not be included as a feature, as it could then be used by the model to make decisions and may introduce unfair bias.

However, simply removing the data from the model does not guarantee fair outcomes; other data in the training data set may be highly correlated with the [MINORITY](#) field. For example, if one of the hotspot locations is close to an area with a high minority population, that correlation could cause the model to be unfairly biased against minorities. This is known as [indirect bias](#). The hotspot feature that is highly correlated with the [MINORITY](#) data is referred to as a [proxy feature](#) for [MINORITY](#).

The watsonx.governance monitoring service can identify proxy features in the training data and detect indirect bias, as long as submissions to the model include [MINORITY](#) as part of the request metadata, which is tracked in the datamart that stores model input and output, but is not actually sent to the model for use in its prediction.

Configuring indirect bias detection is typically done in code using the Python SDK, and is beyond the scope of this lab. For now, you will proceed using a model that has [MINORITY](#) as a feature, to see how the fairness monitor works.

1. From the list of [Evaluations](#) on the left, click on [Fairness](#).

Configure OpenScale evaluation settings

emartens - Policy risk testing

Evaluations

- Fairness **1**
- Quality
- Drift v2
- Drift
- Generative AI Quality
- Model health

Model details

Description

Provide information about the training data and deployed model output to prepare Watson for monitoring and providing explanations for model transactions.

[Reconfigure model](#)

Training data

Storage type
Offline (CSV file)

Training data file name
policy_risk_train.csv

Training data label

Label column
RISK

Training data features

Feature columns
CLAIM_TOTAL, HOTSPOT1, HOTSPOT2, HOTSPOT3, HOTSPOT4, HOTSPOTS5, INITIAL_ODOMETER, LOW_MILEAGE_USE, MAKE_MODEL, MALE25, MINORITY, MODEL_YEAR, NB CLAIMS, NB_DRIVERS,

2. Note the description of fairness in the center of the screen, which gives a good definition of what the monitor is evaluating. Click on the [Edit](#) icon in the [Configuration](#) tile.

Configure OpenScale evaluation settings

emartens - Policy risk testing

Evaluations

- Fairness **1**
- Quality
- Drift v2
- Drift
- Generative AI Quality
- Model health

Fairness

Description

The Fairness monitor checks your deployments for biases. It tracks when the model shows a tendency to provide a favorable (preferable) outcome more often for one group over another. You will specify which values represent favorable outcomes, select the features to monitor for bias (for example, Age or Sex), and specify the groups to monitor for each selected feature.

Configuration

To select a configuration type, click the edit icon. **2**

Favorable outcomes

To select the favorable outcomes, click the edit icon.

3. You will manually configure the fairness file. Leave [Configure manually](#) selected and click [Next](#) to proceed.

To properly monitor a model for unfair bias, you must specify which model outcomes are favorable and which are unfavorable. For binary classification models like a Risk vs. No Risk credit model, or a Hire vs. No Hire candidate screening model, these values are easy to determine. However, it's a bit more difficult for a regression model like the one in this lab. You will need to define ranges of outputs that represent favorable or unfavorable outcomes. Note that the monitoring tool has read in the training data, and helpfully filled in what the minimum and maximum values for **RISK** are in that dataset.

4. For this use case, you will identify any score of 40 or higher as an unfavorable outcome. Use the number entry fields to enter a minimum value of **0** and a maximum value of **39**, then click [Add value](#).

emartens - Policy risk testing

Evaluations

- Fairness **1**
- Quality
- Drift v2
- Drift
- Generative AI Quality
- Model health

Fairness

Select the favorable outcomes

Description

For each group, Watson OpenScale will calculate the percentage of transactions that receive a favorable outcome.

Select the values that represent favorable (preferable) outcomes. You can also add a value manually if it is not included in the list.

Add value ranges and select the favorable outcomes

Minimum value: **0** Maximum value: **59**

4 Add value

Values	Favorable	Unfavorable

5. Repeat previous step to add a second value, with a minimum value of **40** and a maximum value of **100** (the theoretical upper limit of the model output), then click [Add value](#).

6. Use the checkboxes to set the first value to **Favorable** and the second value to **Unfavorable**, then click **Next**.

7. Set the **Minimum sample size** to 100 and click **Next**.

8. Leave the **Selected monitored metrics** set to **Disparate impact** and click **Next**.

9. Leave the lower and upper thresholds for **Disparate impact** set to their defaults, and click **Next**.

You now need to select which fields to monitor for fairness. The monitoring service will analyze the training data and suggest that **PRIM_DRIVER_AGE** and **PRIM_DRIVER_GENDER** be monitored, as based on their names and values they likely represent age and gender fields. Note that this analysis may take up to a minute to perform.

However, for this use case, you will not need to monitor for these fields, as insurance companies have proven over time that male drivers, as well as drivers in certain age groups, present an elevated risk, and

this data can therefore be legally used to set policy premiums.

10. Use the checkboxes to deselect PRIM_DRIVER_AGE and PRIM_DRIVER_GENDER.

11. Scroll to the bottom of the feature list, and check the box next to MINORITY. Click Next.

12. Use the checkboxes to specify MINORITY as the Monitored group and NON-MINORITY as the reference group. Note that occasionally, MINORITY and NON-MINORITY do not appear as possible values. If this occurs, add the values using the Add custom value field, ensuring to use the exact, case-sensitive text. Click on the Next button.

Values	Monitored	Reference	Recommended
MINORITY	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
NON-MINORITY	<input type="checkbox"/>	<input checked="" type="checkbox"/>	

13. Use the default alert threshold (80), and click Save to finish configuring the fairness monitor. It may take up to a minute for the configuration to save, at which point you will be returned to the model settings screen.

Configure the quality monitor

Next, you will configure and set alert thresholds for the quality monitor.

1. From the list of Evaluations on the left, click on Quality.

2. Click the **Edit** icon on the **Quality thresholds** tile.

3. Leave the default lower and upper threshold values as they are. Note that you can click the **Information** icon to the right of each value for more information on how it is calculated. Click **Next**.

4. Set the **Minimum sample size** value to 100. Click **Save** to save the quality configuration.

Configure the explainability service

Next, you will configure the service that provides in-depth explanations for individual model predictions.

1. In the **Explainability** section, click on **General settings**.

2. In the **Explanation method** tile, click on the **Edit** icon.

3. Two different methods are available for explanations: **Shapley Additive Explanations** (SHAP) or **Local Interpretable Model-agnostic Explanations** (LIME). As described in hint that appears when you click

the [Information](#) box, SHAP often provides more thorough explanations, but LIME is faster.

4. Leave the LIME method selected and click [Save](#).

Run an evaluation

Now that the model monitors have been configured, you can run an evaluation of the model.

1. Return to the model screen by clicking on the [X](#) button in the upper right to close the [Configure OpenScale evaluation settings](#) window.
2. Click on the [Actions](#) button to open the [Actions](#) menu.
3. Click on [Evaluate now](#). The [Import test data](#) screen opens.

The screenshot shows the OpenScale interface for the 'emartens - Policy risk testing' model. The top navigation bar includes 'API reference', 'Test', 'Evaluations' (which is underlined), 'Transactions', and 'AI Factsheet'. Below this, there's a summary section with 'Last evaluation: --', 'Test data set: --', and 'Number of explanations: 0'. To the right, there are metrics: 'Tests run: 0', 'Tests passed: 0', and 'Tests failed: 0'. A large circular icon also displays '0 Tests run'. On the far right, a vertical 'Actions' menu is open, containing three items: 'Evaluate now' (with a hand cursor icon, circled with a black circle and labeled '2'), 'Configure monitors', and 'View model information' (circled with a black circle and labeled '3').

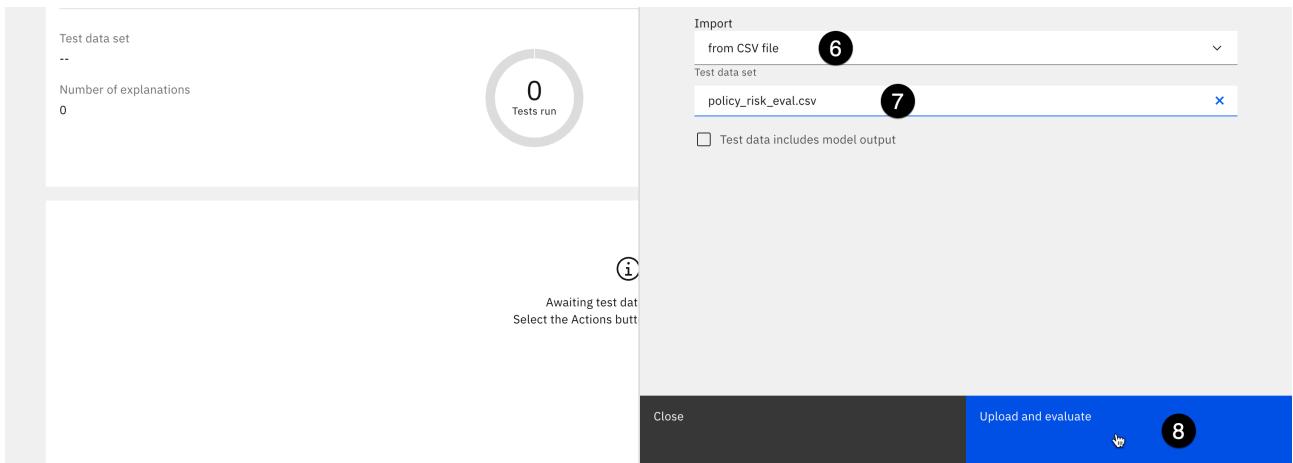
4. Right-click on [this link to the evaluation file](#) and download it to your machine. Note that this file is also available in the project you are using for this lab.

NOTE: When downloading the file, some web browsers may change the extension of the file to [.txt](#), in which case you will need to rename the file to preserve the [.csv](#) extension.

6. Click the [Import](#) dropdown and choose [from CSV file](#) from the list of options.
7. Drag and drop the downloaded CSV file into the designated area on your screen, or browse to it on your machine using the link.
8. Click [Upload and evaluate](#). Note that the evaluation can take up to several minutes to perform. If it fails for any reason, following the same steps and re-running the evaluation typically fixes the issue.

In rare cases where the monitors time out or fail, you may need to remove the subscription by returning to the Insights dashboard, clicking the three dots on the model tile, and then clicking [Remove deployment](#). You can then re-add the deployment using the steps above. Note that the subscription ID that you pasted in the notebook will change, but the API key and evaluation datamart ID will remain the same.

When the monitor completes and the metrics are displayed, proceed to the next step.



View the quality results

Take a moment to review the results of the evaluation. Note that, based on the content of the random sample of the evaluation data, your results will vary each time you perform the evaluation.

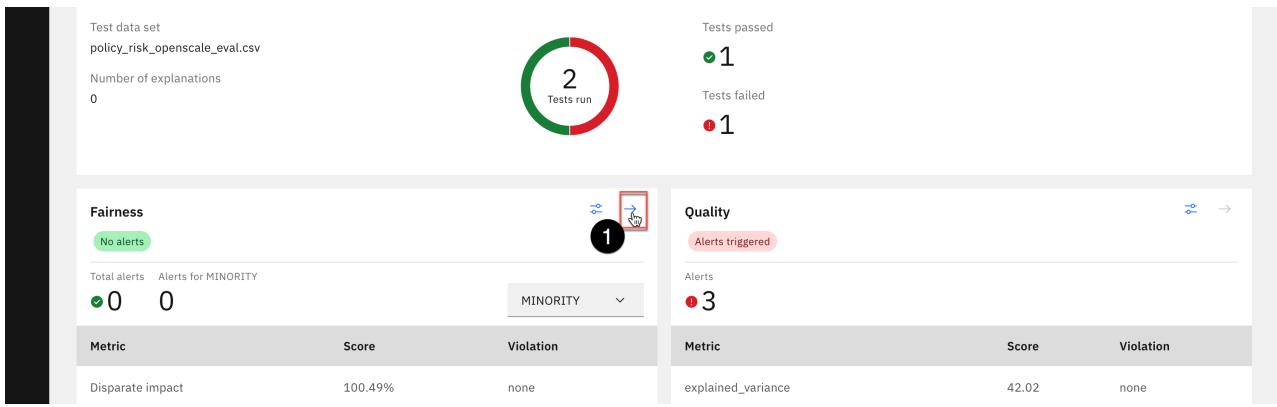
1. Review the different metrics in the quality table. In the screenshot below, notice that, if the measurement falls below the alert threshold set when you configured the quality monitor, the amount will be listed in the **Violation** column of the table. For a full explanation of the many different metrics used to calculate model quality, see this [documentation page](#).

emartens - Policy risk testing Deployed Online

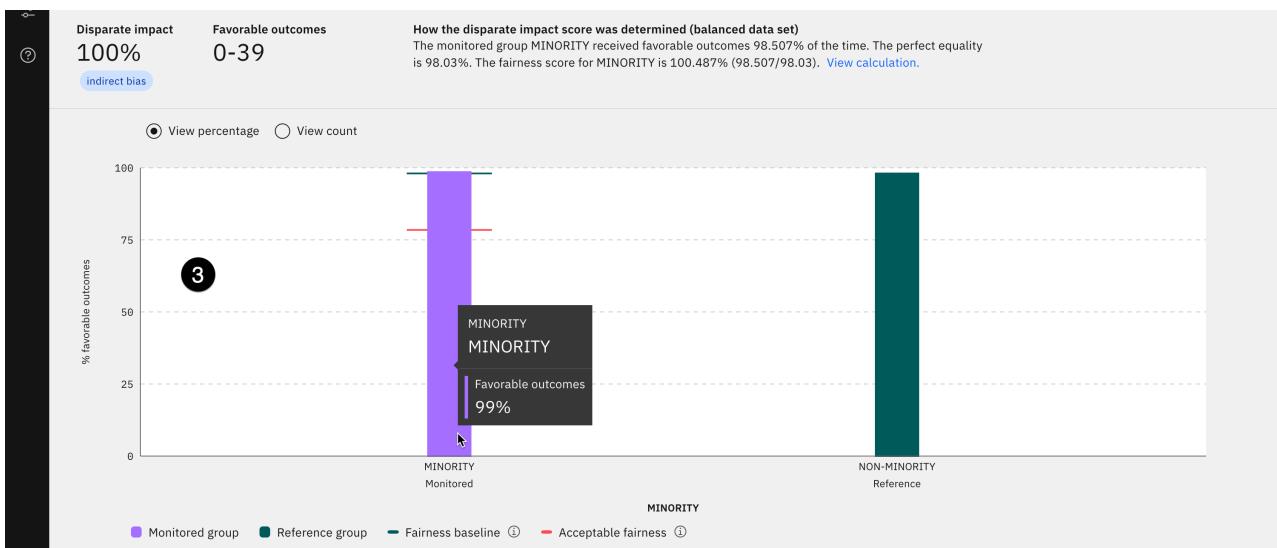
API reference	Test	Evaluations	Transactions	AI Factsheet
Fairness No alerts		Quality Alerts triggered		
Total alerts: 0 Alerts for MINORITY 0 0	MINORITY	Alerts 3		
Metric	Score	Violation		
Disparate impact	100.00%	none		
Associated group: MINORITY				
100 records evaluated				
200 records evaluated				

View the fairness results

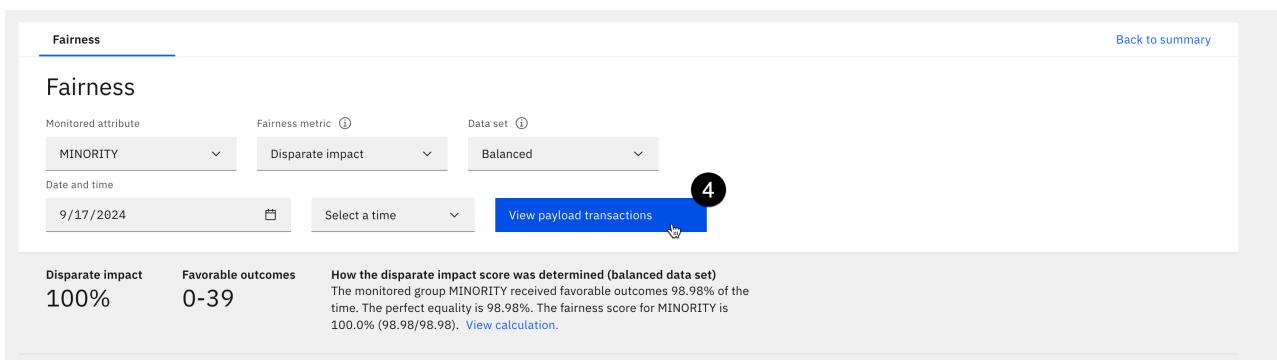
1. Next, look at the **Fairness** tile. Again, based on the content of the random sample of the evaluation data, your results will vary each time you perform the evaluation. In most cases, the model will show as fair, with no alerts for fairness issues. Click on the arrow icon on the **Fairness** tile for more information.



2. Scroll down to the graph portion of the screen and take a moment to read and understand the [How the disparate impact score was determined](#) section, clicking on the [View calculation](#) link to see the specific calculation.
3. Look at the graph. The monitored group, colored purple in the screenshot, has a calculated fairness above the alert threshold (80%, the red line on the graph) that you configured when setting up the fairness monitor. Hovering your cursor over either bar of the graph will also show you the exact percentage of favorable outcomes the group received from the model.



4. When you are finished reviewing the results, scroll to the top of the screen and click on the [View payload transactions](#) button.



8. Explain a prediction

Beyond meeting standards for quality and fairness, AI models in many cases are required to provide explanations into the decisions or predictions they make. For example, under the Equal Credit Opportunity Act in the United States and the European Union General Data Protection Regulation, people affected by an

AI decision have the right to know specific reasons for the decision. The [Right to explanation Wikipedia page](#) provides several useful links with more information.

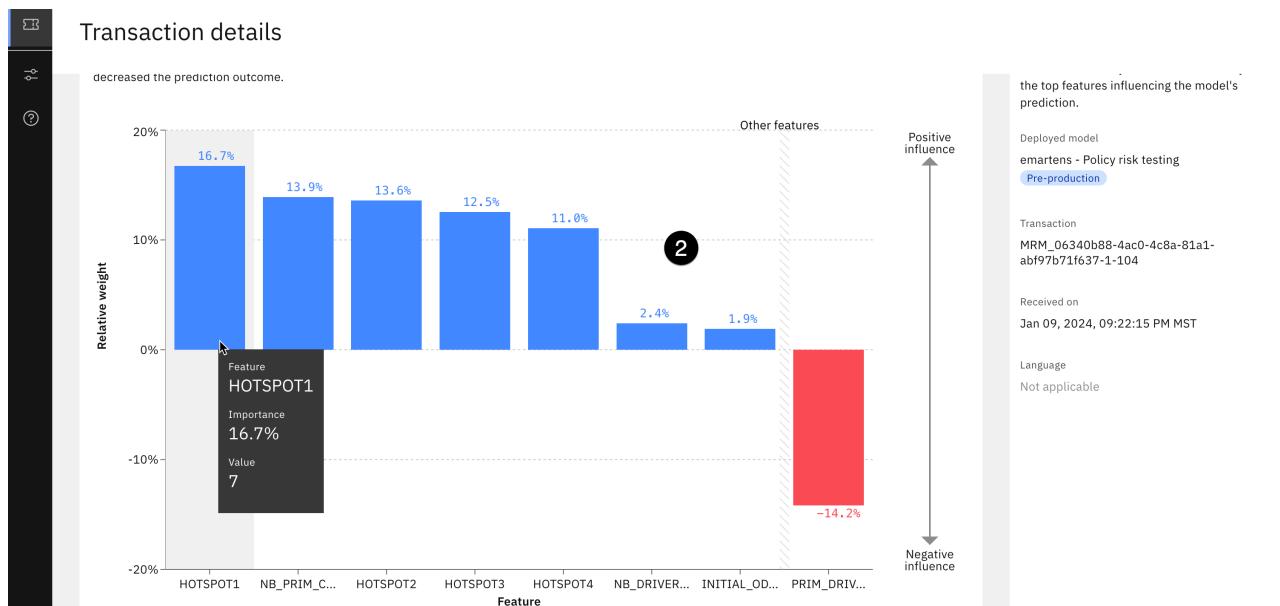
Watsonx.governance provides the ability to generate detailed explanations for predictive models using the algorithm you specified previously when configuring the explainability service.

- From the table of transactions, click one of the [Explain prediction](#) links. You may get more interesting results if you can find a prediction that is close to the threshold for unfavorable (39, as defined when configuring the fairness monitor). The explainability service will use the LIME algorithm to generate a detailed explanation, which can take a few minutes to run.

	Transaction	Group	Subgroup	Group bias	Individual bias	Prediction	Actions
?	MRM_06340b88-4ac0-4c8a-81a1-abf97b71f637-1-1	Reference	NON-MINORITY			17.722501754760742	Explain prediction
?	MRM_06340b88-4ac0-4c8a-81a1-abf97b71f637-1-10	Monitored	MINORITY			23.19902801513672	Explain prediction
?	MRM_06340b88-4ac0-4c8a-81a1-abf97b71f637-1-100	Monitored	MINORITY			23.52501678466797	Explain prediction
?	MRM_06340b88-4ac0-4c8a-81a1-abf97b71f637-1-101	Reference	NON-MINORITY			21.74642562866211	Explain prediction
?	MRM_06340b88-4ac0-4c8a-81a1-abf97b71f637-1-102	Monitored	MINORITY			27.710124969482422	Explain prediction
?	MRM_06340b88-4ac0-4c8a-81a1-abf97b71f637-1-103	Reference	NON-MINORITY			35.64530563354492	Explain prediction
?	MRM_06340b88-4ac0-4c8a-81a1-abf97b71f637-1-104	Reference	NON-MINORITY			35.223262786865234	Explain prediction 1
?	MRM_06340b88-4ac0-4c8a-81a1-abf97b71f637-1-105	Monitored	MINORITY			34.09893035888672	Explain prediction
?	MRM_06340b88-4ac0-4c8a-81a1-abf97b71f637-1-106	Monitored	MINORITY			27.008731842041016	Explain prediction

- Once the explanation has been generated, scroll down to the graph, which shows the influence different features had in the model's outcome. Features in blue increased the final score, while those in red decreased it. For classification models, blue features contributed positively to the model's confidence in the prediction, while those in red decreased the confidence.

Note that your explanation will look different than the screenshot below, which shows the primary contributors to the risk score assigned were proximity to two accident hotspots, as well as the number of claims the primary policyholder has filed. Hover your cursor over the individual columns of the graph for more information.



- Click on the [Inspect](#) tab.

The screenshot shows the 'Explanations' tab for a policy risk testing record. It displays 'Transaction details' and 'Feature influence analyzed with LIME'. The LIME section includes a summary of top features influencing the prediction, a predicted outcome of 35.223262786865, and a count of 234. A note explains how the prediction was determined based on LIME analysis.

On this tab, you can alter values associated with the record and re-submit it to the model to see how the final risk calculation changes. This can be useful for understanding how the model is working, or if a policyholder is looking for ways to decrease their risk assessment.

9. View the updated lifecycle

Having generated metrics for your model, you can now see the updates in the model lifecycle.

1. Return to the IBM watsonx home page. You may need to follow the instructions starting from step 2 in the [Reserving a TechZone shared account](#) section to sign in.
2. Click on the [Navigation menu](#) in the upper left (A) to expand it. Locate the [AI governance](#) section of the menu (B), expanding it if necessary, and click on the [AI use cases](#) menu item (C).

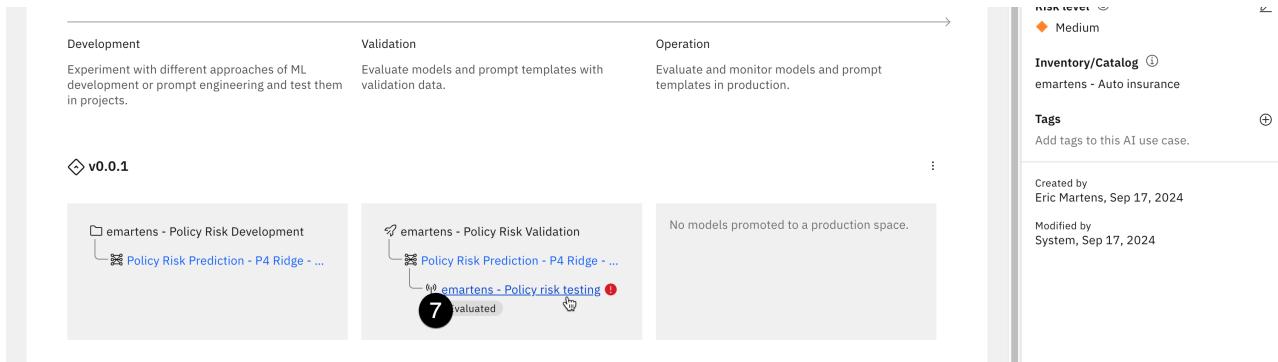
The screenshot shows the 'AI use cases' section of the navigation menu. It displays three main use cases: 'Chat and build prompts with foundation models', 'Build an AI agent to automate tasks with Agent Lab', and 'Tune a foundation model with labeled data with Tuning Studio'. Each use case has a 'Start chatting...' button and an 'Open Prompt Lab' button.

4. Note that, in the table of use cases, an alert is listed in the [Alerts in](#) column, reflecting the quality alerts discovered by the evaluation you ran in a previous step.
5. Click on the auto policy use case you have been using for this section of the lab.

The screenshot shows a table of AI use cases. The first row, 'emartens - Auto policy risk', is highlighted with a red circle containing the number 5. The second row, 'emartens - Claim summarization', is also present. The columns include Name, Status, Owner, Inventory, Tags, Risk level, and Alerts in.

Name	Status	Owner	Inventory	Tags	Risk level	Alerts in
emartens - Auto policy risk	Development in progress	Eric Martens	emartens - Auto insurance		Medium	1 dimension
emartens - Claim summarization	Development in progress	Eric Martens	emartens - Auto insurance		Medium	none

6. Click on the [Lifecycle](#) tab.
7. Scroll down to the lifecycle visualization. Note that the model now shows in the *Validate* section of the lifecycle. A red alert badge appears next to the name of the deployed model, providing a visual cue that the model may have issues. Click on the name of the deployed model. The model deployment's factsheet opens in a new window.



8. Scroll down to the [Quality](#) and [Fairness](#) sections of the model information screen. Note that the evaluation metrics generated by the monitoring tool have automatically been stored on the model's factsheet, allowing stakeholders such as risk managers and data scientists access to information they need to assess model performance, with an optional link provided that will open the monitoring tool if they further information.

10. Next steps

In this section of the lab, you took a model through the develop, test, and validate sections of the lifecycle, and saw how those changes, the metadata, and metrics collected along the way were reflected in the model's factsheet. If you wish, you can continue to progress the model to the *Operate* phase by promoting it to a production deployment space, associating that space with the [Operation](#) phase of the use case, configuring that production deployment for monitoring, and performing metrics evaluations on it. The [watsonx.governance Level 4 PoX lab](#) contains detailed instructions on performing these steps.

If you are using a shared environment, when you are finished recording or delivering any demos, please take a few moments to remove assets you have created. This includes monitoring configurations, deployments, deployment spaces, projects, AI use cases, and model inventories.

When you are ready, you may proceed to the next step and see how watsonx.governance allows organizations to provide similar governance capabilities for generative models as well.

Governing generative models

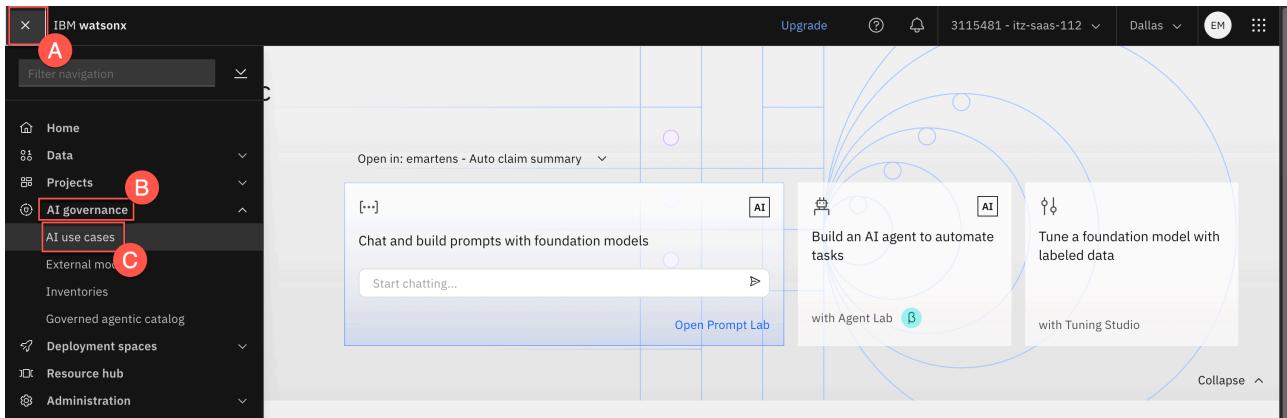
In this section of the lab, you will govern a watsonx.ai foundation model. In this case, the automotive claims department wants to use a generative AI model to summarize auto insurance claims, which they believe will significantly reduce the workload for their claim review department and improve claim response time. However, they have seen news reports about potential issues with generative AI models providing inaccurate information, being prone to responding with hateful, aggressive, or profane (HAP) speech, or potentially leaking personally identifiable information (PII). They need to put a system in place to monitor and evaluate their model to address their concerns before they begin using the model in production.

1. Define an AI use case

As with predictive models, stakeholders in the company will need to create an AI use case to govern their claim summarization model. This AI use case (which is stored in the model inventory identified - or created - in the first step of the lab), will allow administrators, AI engineers, risk managers, and data scientists to organize the different versions, data, and environments used to build and deploy the candidate models.

1. Sign in to IBM watsonx. You may need to follow the instructions starting from step 2 in the [Reserving a TechZone shared account](#) section to sign in.

2. Click on the **Navigation menu** in the upper left (A) to expand it. Locate the **AI governance** section of the menu (B), expanding it if necessary, and click on the **AI use cases** menu item (C).



3. Click the **New AI use case** button to open the **New AI use case** window. Note that if you do not have a **New AI use case** button, you will need to create or gain access to a model inventory, as described in the **Identify a model inventory** section.

Name	Status	Owner	Inventory	Tags	Risk level	Alerts in
Claim summarization	Development in progress	EM Eric Martens	Default Inventory		Medium	None

4. Give your use case a name (A). (For example, "Claim summarization".) If you are using a shared account, add some identifying information such as **your** email address to mark it as belonging to you.

Give your use case a description (B) for the business issue it is attempting to solve.

Use the **Risk level** dropdown (C) to set the associated level of risk. In a real-world example, this would be performed by the risk management officer of an organization. Future integration with IBM OpenPages will offer more risk management functionality.

Use the **Inventory** (D) dropdown to select the model inventory you identified in a previous step.

New AI use case

Create a use case to define a business problem, request a model, and specify details such as risk level and status.

General information

Name (Required) **Claim summarization** (A)

Description **Summarize auto insurance claims** (B)

Risk level **Medium** (C)

Inventory (Required) **Default Inventory (itz-saas-112)** (D)

Owner: EM Eric Martens emartens@us.ibm.com

Status: Draft

Tags: Search tags

5. Click on the **Status** dropdown on the right side of the screen, and note the different values available.

New AI use case

Create a use case to define a business problem, request a model, and specify details such as risk level and status.

General information

Name (Required)
Claim summarization

Description
Summarize auto insurance claims

Risk level
Medium

Inventory (Required)
Default Inventory (itz-saas-112)

Owner
Eric Martens
emartens@us.ibm.com

Status

- Draft
- Draft
- Ready for use case approval
- Use case approved
- Use case rejected
- Awaiting development
- Development in progress

Watsonx.governance allows for organizations to implement formal approval procedures around their model lifecycles. In this case, the auto policy team has requested development of a new model to take into account information gathered by their data scientists. They create a use case for the business problem to request further exploration and development (the *Ready for use case approval* status), and can update the status of the use case as it moves through the process, from initial approval (the *Use case approved* status) to the assigning of AI engineering resources (the *Awaiting development* status). For this lab, you will move directly to an approved use case.

6. Use the dropdown to set the **Status** (A) to **Development in progress**.

Click on the **Create** button (B). After a minute or so, the use case will be created.

General information

Name (Required)
Claim summarization

Description
Summarize auto insurance claims

Risk level
Medium

Inventory (Required)
Default Inventory (itz-saas-112)

Details

Purpose
What's the purpose of this AI use case?

Supporting documentation

Status
Development in progress

Tags

Add tags to this AI use case.

Cancel Create

7. Take a moment to review the use case screen, and note the **Access** tab, which allows sharing of the use case with other stakeholders to allow collaboration on the model lifecycle.

The screenshot shows the IBM Watsonx interface for creating a new AI use case. The top navigation bar includes 'Upgrade', a help icon, a bell icon, '3115481 - itz-saas-112', 'Dallas', and user 'EM'. Below the navigation is a breadcrumb 'AI use cases / Claim summarization' and a header 'About this AI use case'. The main content area has tabs 'Overview' (selected), 'Lifecycle', and 'Access'. A sidebar on the left lists 'Discover next steps' and 'General information' with fields for Name (Claim summarization), Description (Summarize auto insurance claims), Owner (EM), Status (Development in progress, Aug 10, 2025), Risk level (Medium), Inventory (Default Inventory), and Tags.

You have successfully created a use case for an AI model to address a business need. In the next step, you will associate a model with that use case for evaluation.

2. Create a new deployment space

Deployment spaces contain deployable assets, deployments, deployment jobs, associated input and output data, and runtime environments. You can use spaces to deploy various assets and manage your deployments, and control access to the models and data stored there.

1. Click the **Navigation menu** (A) in the upper left corner to open it. Click on the **Deployments** menu item (B) to expand it. Click on the **View all deployment spaces** menu item (C). The **Deployments** screen opens.

The screenshot shows the IBM Watsonx navigation menu expanded. The 'AI governance' section is selected, revealing 'Deployment spaces' (circled in red B) under 'View all deployment spaces' (circled in red C). The 'About this AI use case' panel is visible on the right, showing the same details as the previous screenshot.

2. Click on the **New deployment space** button.

The screenshot shows the 'Deployments' screen. The top navigation bar and sidebar are identical to the previous screenshots. The main content area shows a heading 'Deployments' with '0 spaces' and tabs 'Activity' (selected) and 'Spaces'. A prominent blue button labeled 'New deployment space +' is highlighted with a red box.

3. Give your deployment space a name (A) with identifying information such as **your email address**, and include "testing" to denote this as a space for testing. Provide a description (B) for your space.

Click on the **Deployment stage** dropdown (C), and click on **Testing** from the list. Designating the deployment stage will ensure that the models deployed in this space will appear in the correct phase of the lifecycle map in the AI use case and will use the testing view in the metrics and evaluation screens in later steps.

Create a deployment space

Use a space to collect assets in one place to create, run, and manage deployments

Define details

Name
emartens - Claim summary test

Description (Optional)
Test the candidate claim summary model

Deployment stage
Testing

Tags (optional)

4. Ensure that the **Select storage service** dropdown (A) is correctly set to the object storage service you are using for this lab. Click the **Select watsonx.AI Runtime** dropdown (B), then click on the watsonx.ai Runtime service you are using for this lab.

Click on the **Create** button (C). Space creation can take up to a few minutes. When it is finished, a popup window will appear informing you that the space is ready.

Add tags to make assets easier to find

Storage
cos-110000b3qc

watsonx.ai Runtime (optional)
wml-110000b3qc

Advanced Settings

Create

5. Once the space is created, click the **close** button to close the popup.

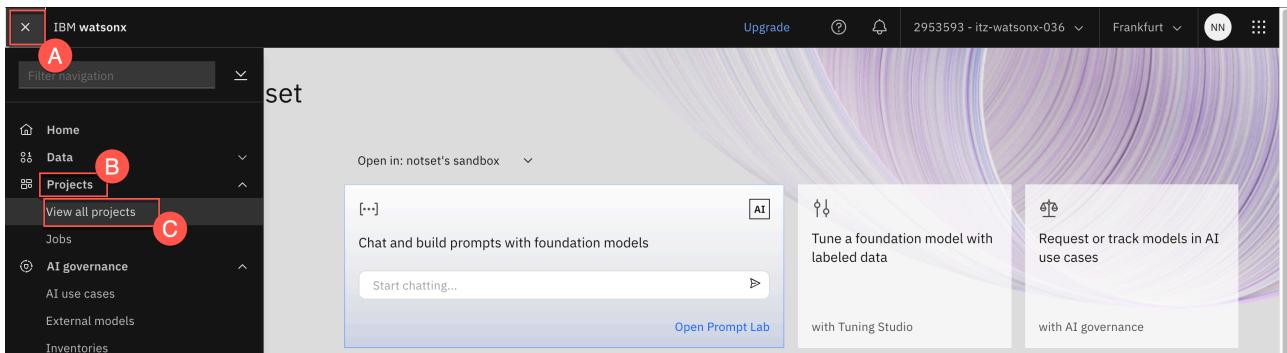
You have successfully created a deployment space for the deployment and testing of your model. Next, you will import assets and start to work with the model.

3. Set up a watsonx.ai project

In this step, you will create an IBM watsonx.ai project that will contain all the assets used to deploy and work with a generative model you will be governing. Watsonx projects provide a central location for data scientists, data engineers, subject-matter experts, and other stakeholders to collaborate on data science projects.

Create the project

1. Right-click link for the **project file** and choose the appropriate menu option for your browser to download it to your machine. Do **NOT** unzip the file.
2. In a separate browser window, navigate to the watsonx projects screen. You can reach the projects screen by click on the **Navigation menu** (A) in the upper left, clicking on the **Projects** item (B) to expand it, then clicking on the **View all projects** menu item (C).

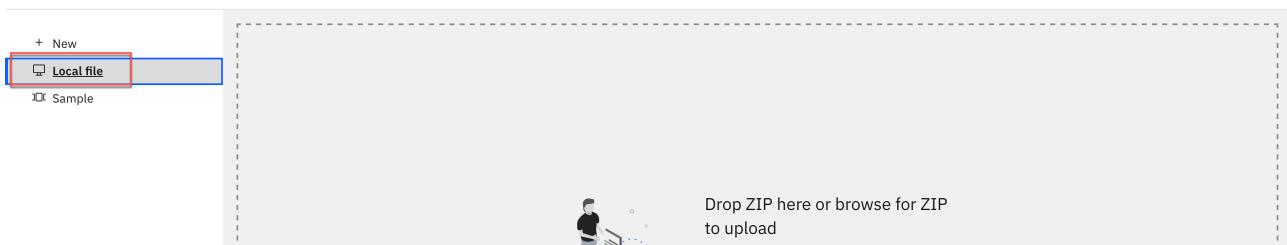


3. Click the blue **New project** button on the right.

4. Click the **Local file** option on the left.

Create a project

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



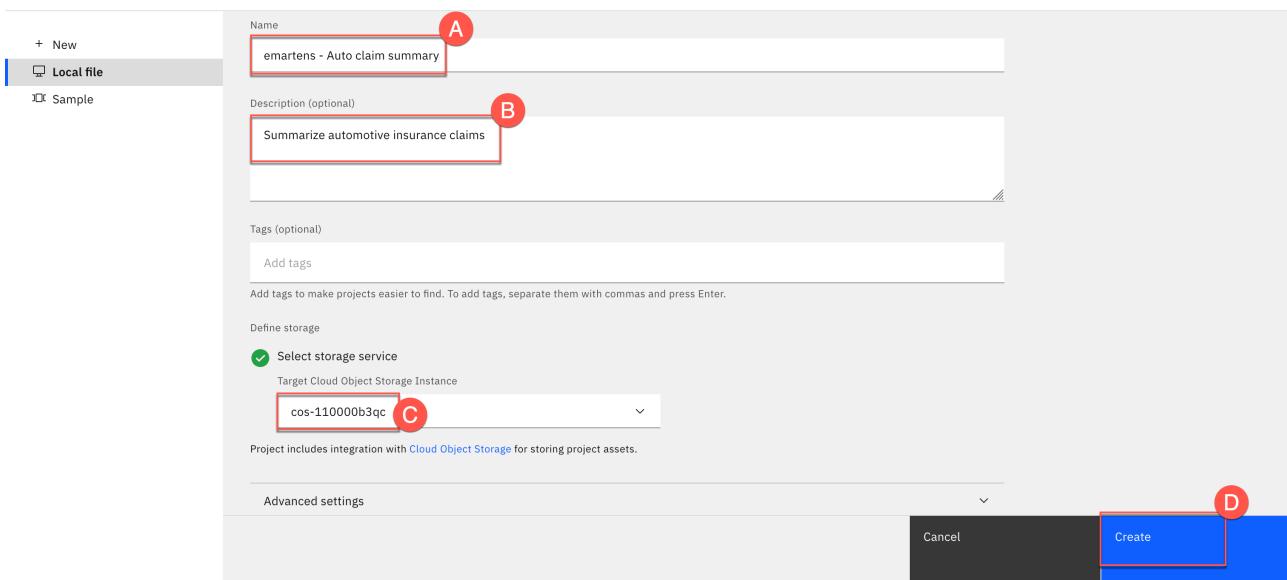
5. Click the **Browse** button in the middle of the screen, and browse to the zipped *Auto-claim-summary.zip* file you downloaded in step one.

6. Give your project a name (A), ensuring that the name begins with some identifying information such as the beginning of your email address or IBM ID. For example, [emartens - Auto claim summary](#).

7. Give your project an optional description (B). Click on the **Select storage service** dropdown (C) and select the Cloud Object Storage instance associated with this environment. Click on the **Create** button (D) to create the project from a file.

Create a project

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



Verify and configure the project

When importing a project from a file, it's critical to ensure that all the resources import successfully, since the project tool will report that creation was successful even if one or more resources failed to import.

1. Click the [View import summary](#) button, and ensure that nothing is listed in the [Incomplete](#) or [Failed](#) categories on the left of the screen. Note that as capabilities change and this lab evolves, your screen may show more assets being imported. The important thing is that all included assets are imported, and that there are no Incomplete or Failed entries.

If an asset failed to import, you will need to return to the projects screen. Locate the project from the list, check the box to the left of it, and then click the [Delete](#) button from the blue menu bar that appears above the table. Then repeat the section above to re-create the project. The vast majority of project import failures can be solved by deleting and re-importing the project.

2. Once the project has successfully imported, click the [Close](#) button.

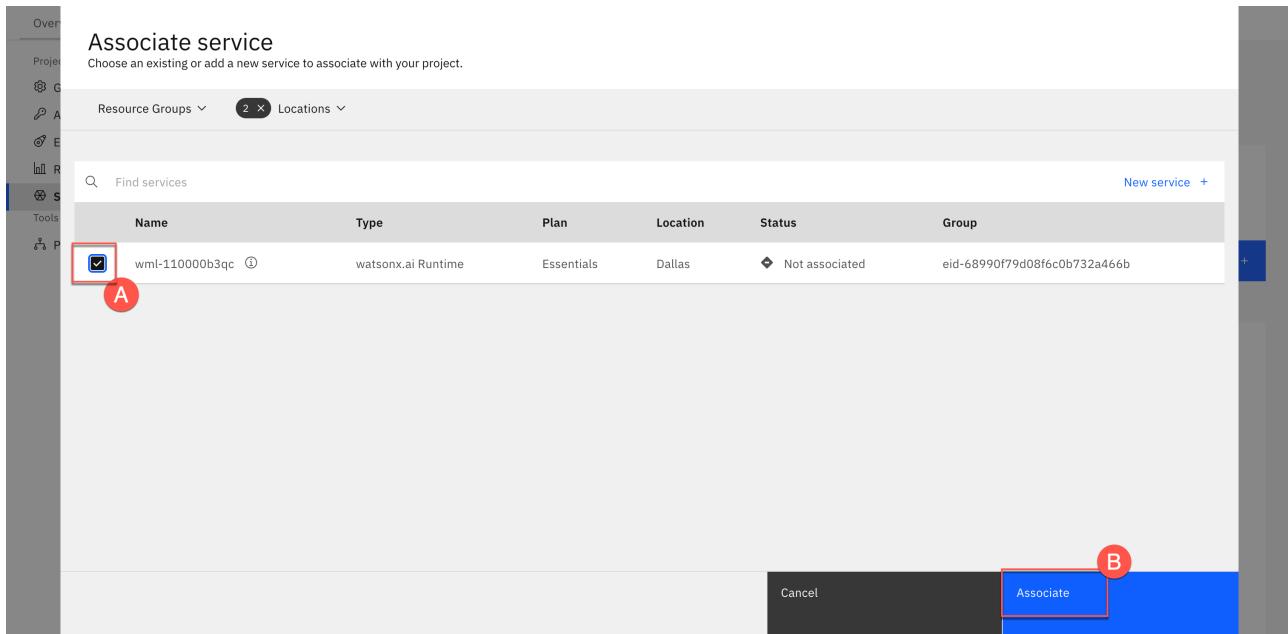
The screenshot shows the 'Project Import Summary' page for a project named 'emartens - Auto claim summary'. The summary table includes columns for 'Import File' (Auto-claim-summary.zip), 'Import Date' (Aug 10, 2025), and 'Imported By' (Eric Martens). A note at the top says 'The project imported successfully'. Below the table is a search bar and a dropdown for 'All asset types'. The main table lists assets categorized by status: 'Imported' (7 items), 'Incomplete' (0 items), and 'Failed' (0 items). The 'Imported' section contains five Data Assets and two Prompt templates. The 'Incomplete' and 'Failed' sections are empty. The table has columns for 'Name', 'Size', and 'Type'.

3. Click on the [Manage](#) tab (A). Click the [Services & integrations](#) item (B) from the menu on the left. Click the blue [Associate service](#) button (C) on the right. The [Associate service](#) popup window opens.

The screenshot shows the 'IBM Watsonx' interface with the 'emartens - Auto claim summary' project selected. The top navigation bar includes 'Upgrade', '3115481 - itz-saas-112', 'Dallas', and other account details. The left sidebar has tabs for 'Overview', 'Assets', 'Jobs', and 'Manage' (which is highlighted with a red circle labeled A). Under 'Manage', the 'Services & integrations' tab is selected (highlighted with a red circle labeled B). The main content area displays the 'Services & integrations' section with tabs for 'IBM services' and 'Third-party integrations'. A callout box provides instructions for associating services. On the right, a 'Find services' search bar and a 'Service type' dropdown are shown. A prominent blue button labeled 'Associate service' with a red circle labeled C is located in the bottom right corner of this section.

6. Locate the appropriate machine learning service for the account in the table. The [Type](#) column should say *watsonx.ai Runtime*, *Watson Machine Learning* or similar. If you have multiple options, choose the one that most resembles the one in the screenshot below. Check the box to the left of the service (A).

Click the blue **Associate** button (B).

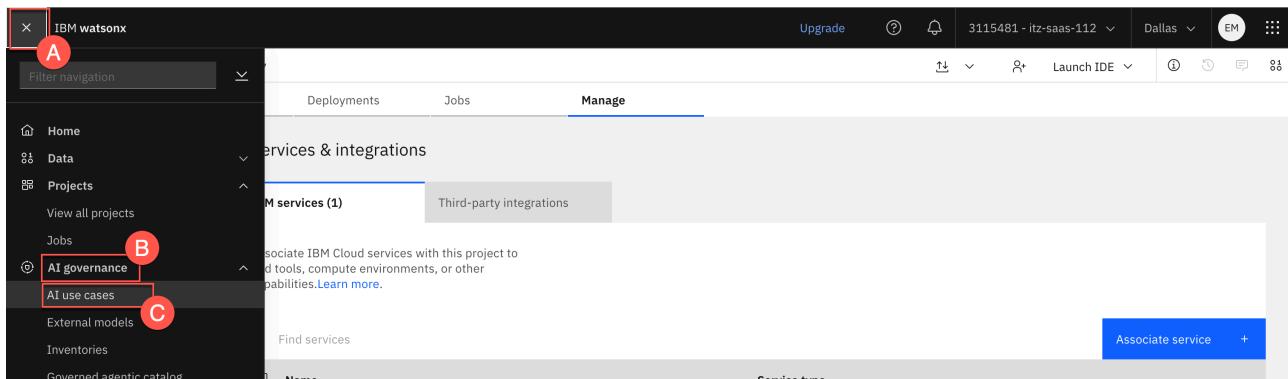


The project is now configured and ready to use.

4. Associate use case workspaces

To enable tracking the model using the use case you created in a previous step, the project and the deployment space must be associated with the different phases of the use case. Projects and deployment spaces can both contain models, assets, metadata, and access controls necessary for AI governance. For this reason, they can be described as **workspaces** for the purposes of the AI use case. Associating workspaces with different phases of the model lifecycle in the use case allows you to organize and manage governance materials in a logical manner.

1. Click on the **Navigation menu** (A) in the upper left corner to expand it. If necessary, click on the **AI governance** menu item (B) to expand it. Click on the **AI use cases** menu item (C).



2. Locate the auto claim summarization use case you created in a previous step and click on it.
3. Scroll down to the **Associated workspaces** section of the screen, and note the three phases of the model lifecycle (*Develop*, *Validate*, and *Operate*). Workspaces (projects and deployment spaces) can be associated with each of the three phases.
4. Click on the **Associate workspace** button in the **Develop** tile. The **Associate workspaces** window opens.

Overview Lifecycle Access

Associated workspaces

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

Development

In this phase your development team will experiment with different approaches of ML development or prompt engineering and test them in projects.

[Associate workspaces](#) +

Validation

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

[Associate workspaces](#) +

Operation

In this phase your development team will evaluate models in pre-production deployment spaces.

[Associate workspaces](#) +

About this AI use case

Name	Claim summarization
Description	Summarize auto insurance claims
Owners	EM
Status	Development in progress(Eric Martens, Aug 10, 2025)
Risk level	Medium
Inventory	Default Inventory
Tags	(+)

If you wish, you can use the [Read more](#) links on this screen to find out about the rules and reasoning behind associating workspaces with phases of the model lifecycle. Note that multiple workspaces can be associated with a single phase of the lifecycle, but a workspace can only be associated with one lifecycle phase. For example, the data science team may be working on several different candidate models for a use case, each with their own assets, collaborators, and datasets. All of these projects would be associated with the *Develop* phase of the single use case. However, each development project would only be associated with the *Develop* phase, and could not be associated with the *Validate* or *Operate* phases.

- From the [Projects](#) section, check the box to the left of the claim summarization project you created in a previous step (A). Click on the [Save](#) button (B) to save the association. The [Associate workspaces](#) window will close, and the project will now appear in the [Develop](#) tile of the [Associated workspaces](#) section of the use case view.

Projects

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

emartens - Auto claim summary

[New project](#) +

Name	Created	Your role	Tracked AI assets	Associates
emartens - Auto claim summary	08/10/2025	Admin	0	0

A

B

- Click on the [Associate workspace](#) button in the [Validate](#) tile. The [Associate workspaces](#) window opens again.

Overview Lifecycle Access

Associated workspaces

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

Development

emartens - Auto claim summary

[Associate workspaces](#) +

Validation

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

[Associate workspaces](#) +

Operation

In this phase your development team will evaluate models in pre-production deployment spaces.

[Associate workspaces](#) +

About this AI use case

Name	Claim summarization
Description	Summarize auto insurance claims
Owners	EM
Status	Development in progress(Eric Martens, Aug 10, 2025)
Risk level	Medium
Inventory	Default Inventory
Tags	Add tags to this AI use case.

- From the **Space** section, check the box to the left of the claim summary deployment space you created in a previous step (A). Click on the **Save** button (B) to save the association. The **Associate workspaces** window will close, and the project will now appear in the **Develop** tile of the **Associated workspaces** section of the use case view.

Name	Created	Your role	Tracked AI assets	Stage	Associates
<input checked="" type="checkbox"/> emartens - Claim summary test	08/10/2025	Admin	0	Testing	0
<input type="checkbox"/> openscale-express-path-preprod-16e15a45-d586-4044-96d9-08b90f9eaf12	08/10/2025	Admin	0		0
<input type="checkbox"/> openscale-express-path-16e15a45-d586-4044-96d9-08b90f9eaf12	08/10/2025	Admin	0		0

Cancel Save

You have successfully created a use case for an AI model to address a business need, and associated your project with it. In the next step, you will begin tracking a model for evaluation.

5. Track the foundation model

This lab will use a generative model trained to summarize insurance claims. If you wish, you can click on the **Insurance claim summarization** from the list of project assets to work with it in the watsonx.ai prompt lab, though this lab will focus strictly on governing the model and not on refining it with the prompt lab.

Configure model tracking

- Return to the list of watsonx projects in your browser. You can reach the projects screen by click on the **Navigation menu** (A) in the upper left, clicking on the **Projects** item (B) to expand it, then clicking on the **View all projects** menu item (C).

- A IBM watsonx
- Filter navigation
- Home
- Data
- Projects View all projects
- Jobs
- AI governance
- AI use cases
- External models
- Inventories

Open in: notset's sandbox

[...]

Chat and build prompts with foundation models

Start chatting...

Open Prompt Lab

↑ ↓

Tune a foundation model with labeled data

with Tuning Studio

☰

Request or track models in AI use cases

with AI governance

- Locate the **Auto claim summary** project from the list and click on it to open it.
- Click on the **Assets** tab of the project.
- From the list of assets, locate the **Insurance claim summarization** entry and click on the three dots to the right to expand the context menu.
- Click on **Go to AI factsheet** from the menu.

The screenshot shows the Watsonx Assets interface. On the left, there's a sidebar with '7 assets' and 'Asset types' sections. The main area lists several AI assets, each with a name, last modified time, and a 'More' options menu. A context menu is open over the first asset in the list, with the 'Evaluate' option highlighted.

6. Take a moment to review the information gathered on this screen, including data on the foundation model it is built on, the prompt templating and basic task, and parameters. As regulations around the use of AI models increase, the ability to automatically track and easily retrieve model metadata without any manual effort from data scientists has become critical. Watsonx provides this capability, improving transparency and helping speed time-to-value for AI models.

7. Scroll back up to the **Governance** section and click on the **Track in AI use case** button.

The screenshot shows the AI Factsheet for the 'Insurance claim summarization' asset. The 'Governance' tab is active. A yellow warning box appears stating: 'This model was automatically untracked due to an association conflict. A project can only be associated to one phase of one AI use case. Importing this asset would've caused a conflict and as a result, your asset has been untracked automatically. You can track it again based on the association of this workspace.' Below this, a note says 'This prompt template is not tracked.' and provides instructions to 'Track in AI use case'.

8. When asked to define an approach, leave **Default** approach selected and click **Next**.

9. When asked to assign a model version, leave **Experimental** selected. Note that you can manually assign a version number here, or choose a more production-ready version number depending on the state of the model. Click **Next** to proceed to the **Review** screen.

10. Click **Track asset** to start tracking the model. The factsheet will reload.

View the updated AI use case

1. Note that the **Governance** section of the model screen now contains information on the approach used, model version, and lifecycle phase (*Develop*).
2. In your browser, navigate to the list of AI use cases by clicking on the hamburger menu in the upper left, then clicking on **AI use cases** from the **AI governance** submenu.

The screenshot shows the IBM Watsonx interface. In the top-left corner, there's a navigation bar with 'IBM Watsonx' and other options like 'Upgrade', '1375031 - Eric Martens's ...', 'Dallas', and a user icon. On the left, a sidebar has 'AI use cases' selected (marked with a red circle '2'). The main area displays four cards: 'Experiment with foundation models and build prompts' (with Prompt Lab), 'Tune a foundation model with labeled data' (with Tuning Studio), 'Request or track models in AI use cases' (with AI governance), and 'Work with data and models in Python or R notebooks' (with Jupyter notebook editor). Below this is a table titled 'AI use cases' with two rows:

Name	Status	Owner	Inventory	Tags	Risk level	Alerts in
emartens - Claim summarization	Development in progress	EM Eric Martens	emartens - Auto insurance		Medium	none
emartens - Auto policy risk	Development in progress	EM Eric Martens	emartens - Auto insurance		Medium	1 dimension

3. From the table, click on the use case you created for this section of the lab.

This screenshot shows the same 'AI use cases' table as the previous one, but with a callout box labeled '3' pointing to the first row where 'emartens - Claim summarization' is listed. The rest of the interface is identical to the first screenshot.

4. Click on the [Lifecycle](#) tab, then scroll down to the lifecycle visualization. Note that the [Insurance claim summarization](#) model is listed in the *Develop* section of the lifecycle.

6. Deploy the foundation model

Next, you will promote the model to a deployment space and deploy it. Watsonx uses deployment spaces to organize models and model-related assets for validation and production access.

Promote the model to the space

1. Return to the watsonx projects screen. You can reach the projects screen by click on the [hamburger menu](#) in the upper left, clicking on the [Projects](#) item to expand it, then clicking on the [View all projects](#) menu item.

This screenshot shows the 'Projects' screen in Watsonx. A callout box 'A' highlights the window title 'IBM Watsonx'. A callout box 'B' highlights the 'Projects' menu item in the sidebar. A callout box 'C' highlights the 'View all projects' menu item under the 'Projects' heading. The main area shows a card for 'Chat and build prompts with foundation models' (with Open Prompt Lab) and another card for 'Tune a foundation model with labeled data' (with Tuning Studio).

2. Click on the name of the project you are using for this section of the lab to open it.

3. Click on the [Assets](#) tab.

4. Locate the [Insurance claim summarization](#) prompt template from the list of assets. Note that it has an icon to the right of the name, denoting that it is being tracked by watsonx.governance and can no longer be edited. Click the three buttons to the right of the asset name to open the context menu.

5. Click on Promote to space. The Promote to space window opens.

The screenshot shows the IBM WatsonX interface with the 'Assets' tab selected (marked with a black circle). A context menu is open over a list of assets, specifically highlighting the 'Promote to space' option (marked with a black circle).

6. Click on the Target space dropdown, and click on the claim summary space you created in the previous step.

7. Check the box to the left of Go to the space after promoting the prompt template.

8. Click the Promote button.

The screenshot shows the 'Promote to space' dialog box. The 'Target space' dropdown is set to 'emartens - Claim summary testing' (marked with a black circle). The 'Go to the space after promoting the prompt template' checkbox is checked (marked with a black circle). The 'Promote' button is highlighted with a black circle.

Create a deployment

The model has been promoted to the space, but cannot be accessed by application developers until it has been deployed.

1. Click the New deployment button. The Create a deployment window opens.

2. Give your deployed model a name with personally identifiable information.
3. Click the **Create** button.

View the deployment details

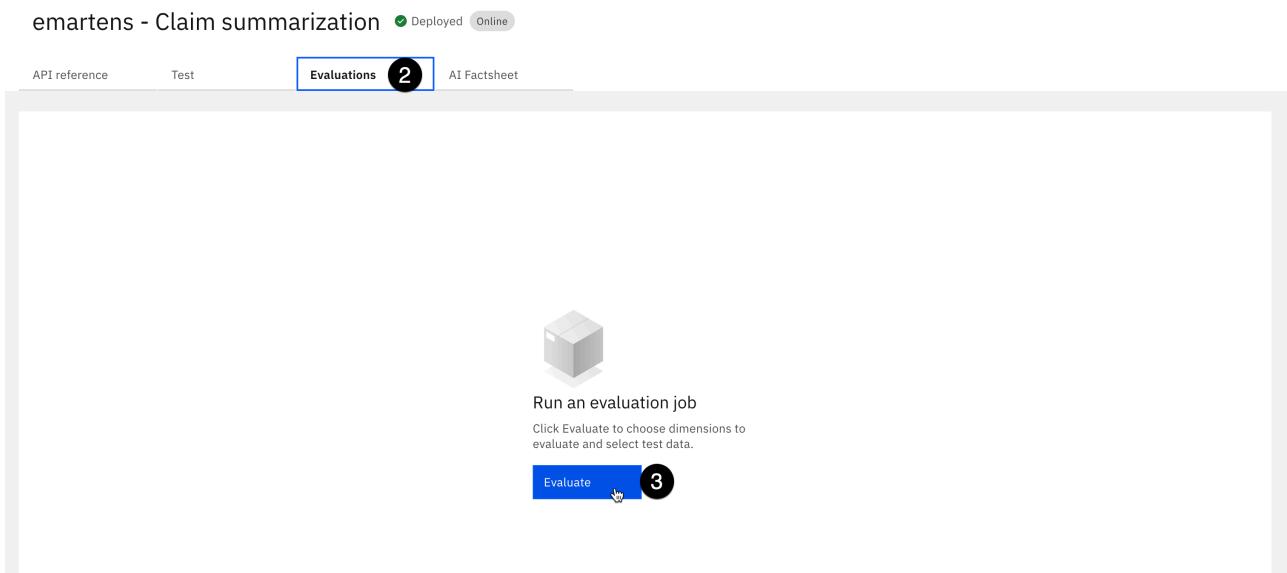
1. Click on the new deployment from the list of deployments.

2. Note that the **API reference** tab contains useful information for application developers looking to integrate the model into their apps, including code snippets and endpoint links.
3. Click on the **AI Factsheet** tab.
4. Note that the tab shows information on the deployment, including scoring URLs. At this time, if you wish, you can return to your list of AI use cases and view this model's use case, which will reflect the same information.

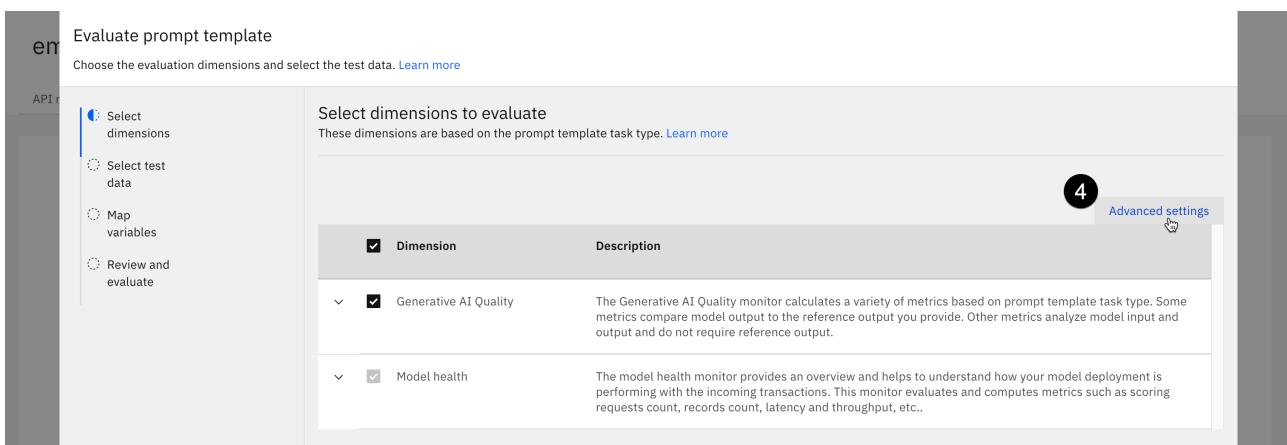
7. Evaluate the foundation model

1. Right-click link for the **evaluation data file** and choose the appropriate menu option for your browser to download it to your machine. **NOTE:** depending on your browser, you may need to rename the file to change the extension from **.txt** to **.csv** after downloading it, as the evaluation service will only accept comma-separated value (CSV) files as input.
2. Click on the **Evaluations** tab of the deployment information screen.

3. Click the **Evaluate** button to open the **Evaluate prompt template** window.



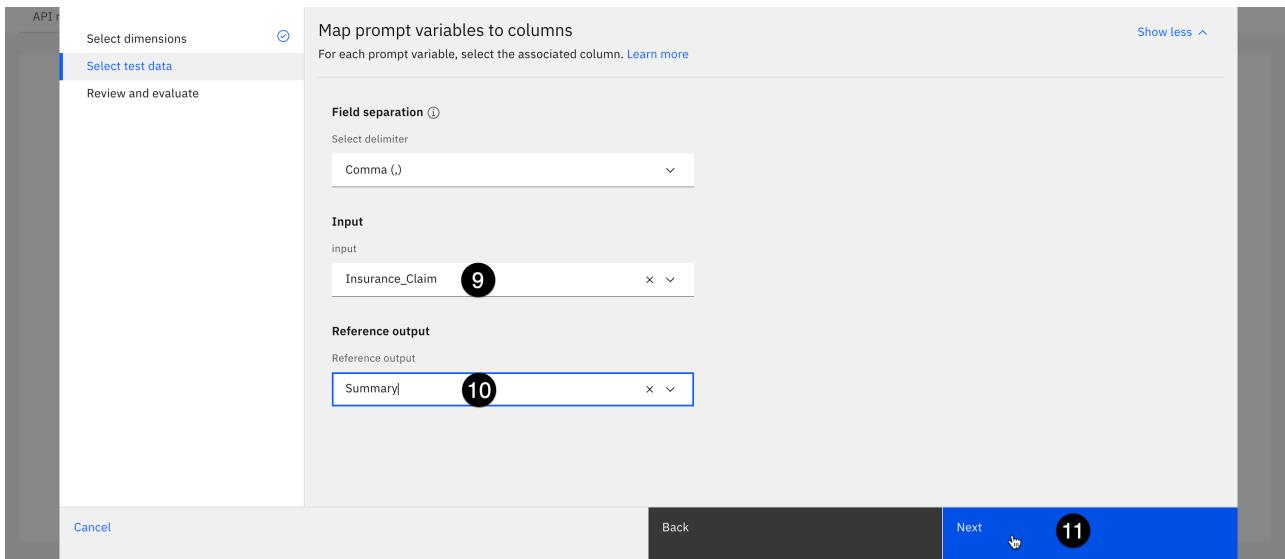
4. The **Select dimensions to evaluate** section of the window shows the different evaluations available. At the time of writing, **Generative AI Quality** and **Model health** are available for this particular prompt template. Click on the **Advanced settings** link.



5. Take a moment to scroll through the **Generative AI Quality** settings screen to see the different metrics that will be measured as part of the quality evaluation, and the alert thresholds set for each. Note that these thresholds can be fully-customized on a per-model basis, allowing risk managers to make sure their models comply with regulatory standards. For more information on the individual metrics, see the [watsonx.governance documentation](#).

The metrics include quality measurements such as precision, recall, and similarity, as well as personally identifiable information (PII) and hateful, aggressive, and profane (HAP) content detection for both model input and output.

6. Click **Cancel** to return to the **Evaluate prompt template** window.
7. Click **Next**.
8. Drag and drop the *claim_summarization_validation.csv* file you downloaded earlier to the upload section on the screen, or browse to it.
9. Click on the **Input** dropdown, and click on **Insurance_Claim** from the list.
10. Click on the **Reference output** dropdown, and click on **Summary** from the list.
11. Click **Next**.



12. Click [Evaluate](#) to start the evaluation, which can take up to a few minutes to run. Note that if the evaluation fails, re-running it will usually complete successfully.

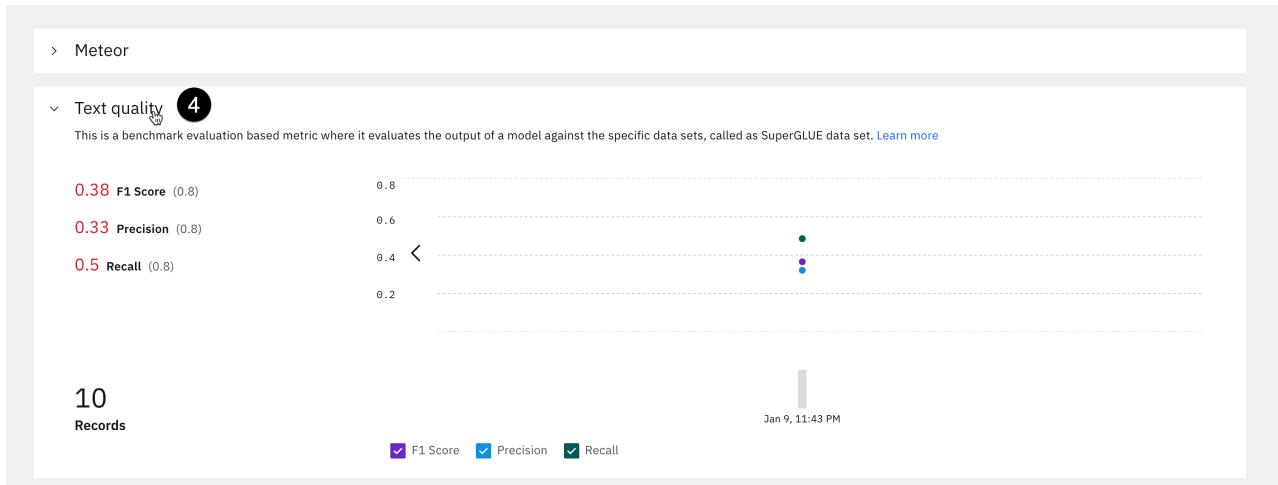
Review the evaluation results

The evaluation summary screen shows the results of the most recent evaluation.

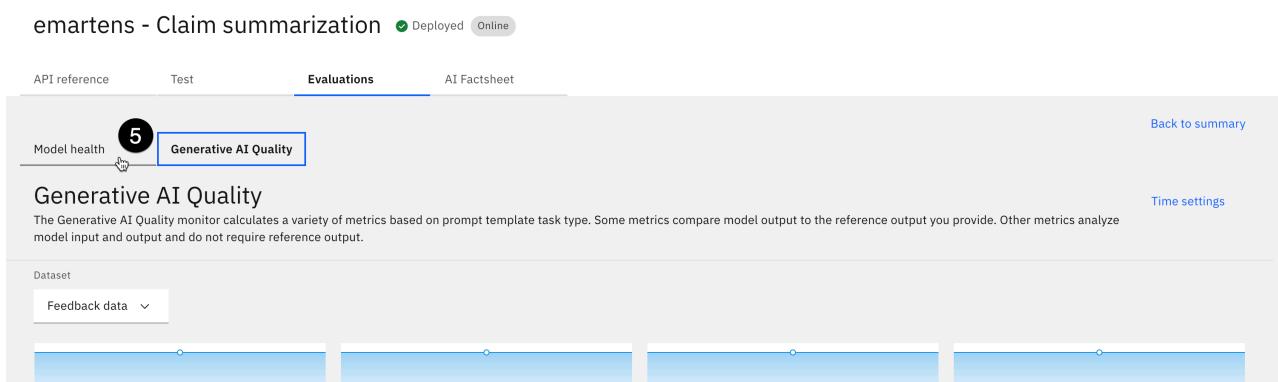
1. Scroll down to the [Generative AI Quality - Text summarization](#) section. The different quality metrics are listed here, with the model's score and any alert threshold violations.
2. Click the [arrow icon](#) for more information on the quality metrics.

Metric	Score	Violation
ROUGE-1	0.41	0.39
ROUGE-L	0.30	0.54

3. 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.
4. Scroll down to the sections for the different metrics. Note that you can click to expand the sections for a more detailed view of each metric.



- When you are finished viewing the quality metrics, scroll back to the top of the screen and click on the [Model health](#) tab.



- Take a moment to review this tab, which contains historical data for health metrics such as latency, throughput, 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 in a reasonable amount of time, and keeping compute costs to acceptable levels.

8. View the updated lifecycle

- Click on the [AI Factsheet](#) tab, which will open the factsheet specific to the model deployment.



- Note that the results of the evaluation have been automatically saved to the factsheet.
- Scroll down to the [Evaluation results](#) section of the factsheet. The information from the model evaluation has been automatically stored in the factsheet, allowing stakeholders such as risk managers, business users, and AI engineers to access relevant information without requiring any manual effort from data scientists.
- Scroll to the bottom of the screen and click on the [More details](#) icon.

The screenshot shows a detailed view of a deployed AI model. At the top, there are performance metrics: Minimum record throughput (0.35), Maximum output token count (53), and Average records (1). Below these, a section titled 'Interested in more details?' contains a note that this information is part of an AI factsheet. A 'More details' button is highlighted with a red box and the number '4'. To the right, deployment metadata is listed: Date (Jan 9, 2024, 11:07 PM), Last updated (Jan 9, 2024, 11:44 PM), Deployment ID (89a3b603-fcf9-46de-8615-7cce10...), Copies (1), Serving name (No serving name), Description (No description provided), and Tags (Add tags to make assets easier to find).

5. The full factsheet for the base model opens, containing all the previous model metadata, as well as the metrics from the deployed version.

Conclusion

Over the course of this lab, you followed the model lifecycle, from initial use case conception through development, to deployment and validation. You saw how the watsonx.governance platform provides a single solution for managing both traditional predictive models and newer generative AI models, tracking model metadata and performance across every phase of the lifecycle. You explored the ways in which watsonx.governance automated tasks like metrics gathering and report generation, and seamlessly provided that information in a fully managed environment so stakeholders can make decisions on the most up-to-date 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, and primarily affects the shared TechZone account with student IDs recommended for most users. The root cause is that the student ID used to log in loses access to the watsonx.governance services before the standard authentication timeout.

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. No access to services / services not appearing

This encompasses a wide variety of issues, including but not limited to:

- [AI Governance](#) not appearing in the watsonx menu
- Watson Machine Learning instances not appearing in the list of available services when configuring a project
- No access to the watsonx.governance monitoring (OpenScale) monitoring dashboard

By far the most frequent cause of these issues is differences in the account, service, and project region. The first step is to clear your browser cache or use private or incognito browsing mode to ensure that your location has not been incorrectly cached by IBM Cloud.

The next step is to verify the region of your TechZone reservation or market account reservation. For TechZone reservations, log into your [My Reservations](#) page on TechZone and click on the tile for your account reservation.

This lab **HIGHLY** recommends using the Frankfurt region for all reservations.

Scroll down to the [Environment](#) section and check the [Region](#) for your account.

Note: Optimized by [IBM Turbonomic](#)

Reservation ID: 660f01f4773d94001e53e783

Type: IBM Cloud

Request method: watsonx-ai-saas

Cloud Account: ITZ-WATSONX

Region: us-south (highlighted with a red box)

Customer data: false

Idle runtime limit:

Transaction ID: 93bf33dc-59ae-40ec-85cd-21ad7477e71b

Geo: americas

Datacenter: dal10

Environment: watsonxai-v0smzye0

Timeout action:

Next, sign in to IBM watsonx using the link: [Americas](#) | [Europe](#) | [Asia Pacific](#).

Verify that the region for watsonx matches the region for your account:

Welcome back, Eric

Train, deploy, validate, and govern AI models responsibly.

Open in: emartens - Auto policy risk

Dashboards

Experiment with foundation models and build prompts

Tune a foundation model with labeled data

Request or track models in AI use cases

Work with data and models in Python or R notebooks

Region: Dallas (highlighted with a red box)

Finally, return to your watsonx project and verify that its region also matches your account and watsonx region:

Projects / emartens - Auto claim summary

Overview Assets Jobs Manage

Project General Access control Environments Resource usage

Access control

Collaborators (2) Access tokens

Find collaborators Add collaborators

Region: US (Dallas)us-south (highlighted with a red box)

As you switch between regions, it is **highly likely** that you will be affected by cache issues. Therefore, please verify that all regions are correct and your cache has been cleared before contacting support with this issue.

If you are using a TechZone shared account with student IDs, your login may have expired. You will likely need to log out, then log back in by following the instructions starting from step 2 in the [Reserving a TechZone shared account](#) section to sign in. In some rare cases, you may even need to clear your browser's cache or use a private/incognito browser window.

3. Compute capacity limits exceeded

This error is most commonly caused when you attempt to provision and use your own instance of a service such as Watson Machine Learning or watsonx.governance (OpenScale). The lab requires more compute hours than are available in free lite plans for services. For this reason, [do not attempt to use your personal services in conjunction with the TechZone accounts, or create new service in the TechZone accounts](#). If you

ignored the instructions and provisioned a service because you did not have access to one, please see the previous section on troubleshooting that issue.

4. 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.

5. Deployment monitoring or policy capacity limits exceeded

IBM Cloud and the [watsonx.governance](#) monitoring service (OpenScale) have hard limits on the number of projects, subscriptions, and policies that can be created. Though TechZone does load balance reservations, using shared accounts with others can occasionally cause these limits to be exceeded.

If you attempt to create a project and receive an error related to policy capacity limits, or if you get an error when attempting to add a machine learning provider or model subscription from the [watsonx.governance Insights dashboard](#), your shared account may have run out of these resources. You can do your part to prevent this by removing projects associated with your account when you have finished the lab. To correct this issue, reach out to [Eric Martens via Slack](#) or use the TechZone support options.

6. Internal server error when accessing the monitoring service

When logging into the [monitoring service](#), you may receive an [Internal server error](#) in the page that loads. Clearing your browser cache can occasionally fix this, but more often you will need to use the direct link to the service. Direct links vary based on which account you are using.

Sign into the [IBM Cloud Resources](#) page. Verify that you are switched to the correct account in the upper right. Click on the [AI / Machine Learning](#) section of the resource list to expand it. Then locate the Watson OpenScale / [watsonx.governance](#) instance from the list, and click on it.

IBM Cloud Search resources and products... Catalog Manage 2695539 - itz-watsonx-14

Resource list

Name	Group	Location	Product	Status	Tags
Filter by name or IP address...	Filter by group or org...	Filter...	Filter...	Filter...	Filter...
Compute (0)					
Containers (1+)					
Networking (0)					
Storage (1+)					
Converged infrastructure (0)					
Enterprise applications (0)					
AI / Machine Learning (5)					
IBM Knowledge Catalog BNA		Frankfurt	IBM Knowledge Catalog	Active	-
Watson Knowledge Catalog-itz	watsonx	Dallas	IBM Knowledge Catalog	Active	-
Watson Machine Learning-itz	watsonx	Dallas	Watson Machine Learning	Active	-
Watson OpenScale-itz	watsonx	Dallas	watsonx.governance	Active	-
Watson OpenScale-itz	watsonx	Dallas	Watson Studio	Active	-
Analytics (1+)					
Blockchain (0)					

From the watsonx.governance information screen, click on the **Launch Watson OpenScale** button to directly launch the service.

Manage Plan

watsonx.governance

Accelerate responsibility, transparency and explainability in your data and AI workflows with watsonx.governance.

Watson OpenScale is a platform that helps organizations manage, monitor, and direct their AI activities.

A Launch in watsonx.governance

B Launch in Watson OpenScale