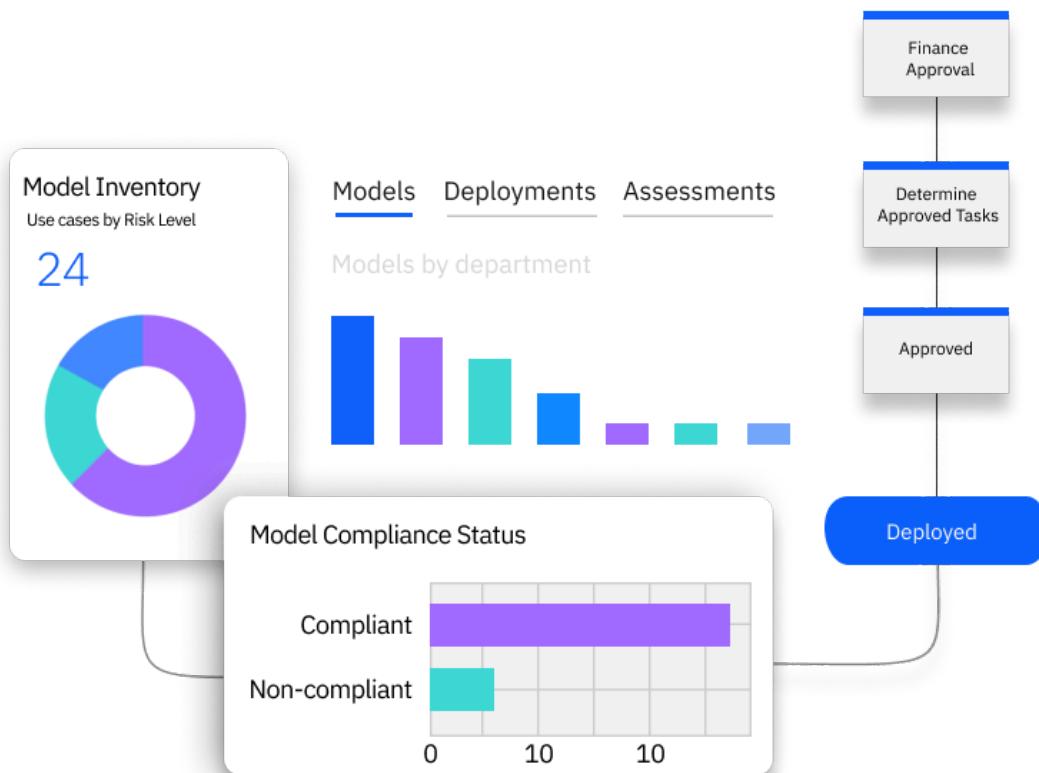


# Monitor and Govern AI with watsonx.governance Day 1

## Hands-on Lab Guide



Eric Martens  
emartens@ibm.com

# Contents

1. Introduction
2. Intake and risk assessment
  1. Log into the environment
  2. Switch contexts
  3. Launch the governance console
  4. Change user profiles
  5. Create a model use case
  6. Progress the use case to the next phase
  7. Identify use case risks
  8. Assess individual risks
  9. Assess multiple risks
  10. Assess applicability for the use case
  11. Approve the use case for development
3. Controls identification and implementation
  1. View the controls the inventory
  2. See controls associated with risks
4. Monitoring and issue management
  1. Explore the development environment
  2. Examine metrics from the monitors
5. Dashboards and reporting
  1. View metrics for the credit risk use case
  2. View metrics for generative AI
  3. View other use cases
6. Conclusion

# Day 1

## Introduction

Welcome to part one of the Deloitte watsonx.governance hands-on lab. In this lab, you will explore the watsonx.governance solution from the perspective of a day-to-day user as the govern a model, from the initial use case request through to production deployment.

In this lab, you will:

- Create an AI use case to address a business problem
- Progress that use case request through a workflow
- Identify risks associated with the use case
- Evaluate the identified risks
- Explore dashboards and reporting on an enterprise level
- Drill down into specific metrics gathered for individual models

## A note on the environment

To complete this lab, you will use a Cloud Pak for Data software environment provisioned on IBM TechZone, with the watsonx.governance software installed. Your lab instructor will provide you with credentials to log into this environment.

Please note that this is a [shared](#) environment, meaning that others in the lab will be able to view as well as edit or delete the artifacts you create. In a production deployment of watsonx.governance, the environment and user accounts would be created and configured to prevent these conflicts; however, that is beyond the scope of this lab. Please be mindful of this, and adhere to the following guidelines:

- Only work with artefacts (use cases, risk assessments) that you or your group has created.
- Include unique identifying information such as your name email address when creating use cases.

## Customization

It is important to understand that [all](#) of the pieces of watsonx.governance that you will use in the following lab are fully configurable. AI use cases can be customized to include any level of detail the client requires. Workflows can be modified or built from scratch to reflect the client's approval processes. Relevant risks and the questionnaires that associate them with use cases can also be completely configured to meet a client's needs.

## Troubleshooting

The environment you are using for this lab has been provisioned on IBM TechZone. Due to infrastructure constraints, you may experience some delay or slow performance. This frequently appears as error messages when saving changes to artefacts such as use cases or workflows. [The vast majority of the time, the save was actually successful](#). Refreshing the screen will fix most of the issues; if this does not work, try the action again. If the action continues to fail, contact your lab instructor for more help.

# Intake and risk assessment

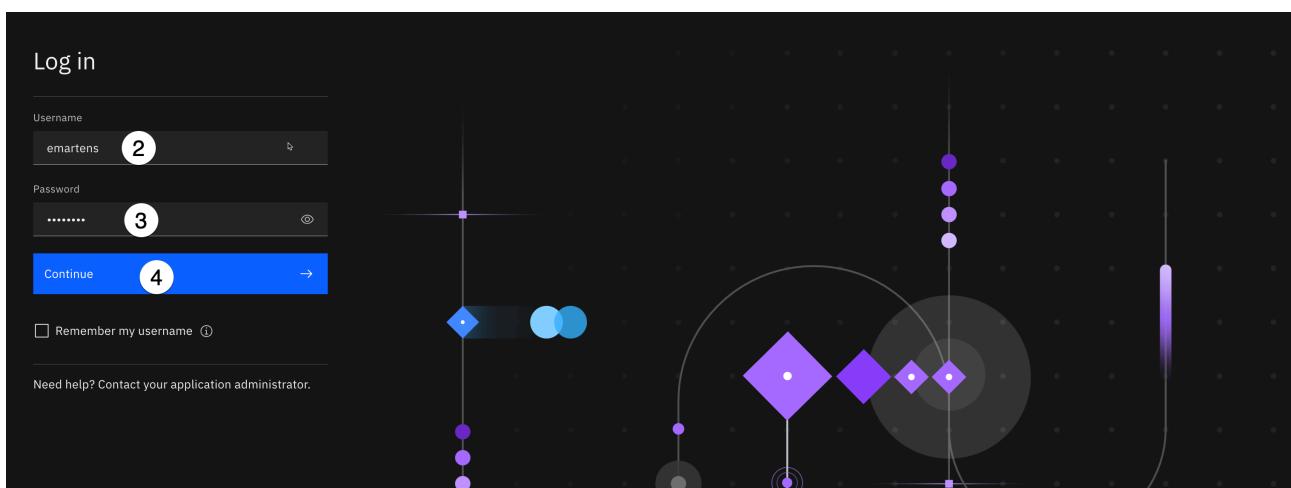
The model governance process begins with the creation of a model use case. A use case is meant to track and capture information about a collection of models and prompts that will be built to serve a particular purpose. A use case should be created whenever there is a business need requiring the use of a model (AI or non-AI) to be built. Model records should then be added as a child of the use case.

In this section of the lab, you will log into the watsonx environment. You will then use the governance console to create a use case, identify associated risks, evaluate those risks, and then review and approve the use case for active development. In completing the actions described in the lab, you will take on many different organizational roles, from the business users making the initial request to the risk managers evaluating the risks. Actually switching between different user personas is beyond the scope of this lab. However, note that watsonx.governance provides fully-featured role-based access control, allowing clients to completely customize user permissions and data access.

## 1. Log into the environment

Your lab instructor will provide credentials for the Cloud Pak for Data or watsonx environment, including a URL, username, and password. Note that in a real-world situation, administrators would likely integrate the environment with whatever authentication method they use in day-to-day operations, and assign permissions, roles, and groups based on those IDs.

1. In a different browser window, navigate to the provided URL.
2. Enter the provided username in the **Username** field.
3. Enter the provided password in the **Password** field.
4. Click on the **Continue** button to log in.

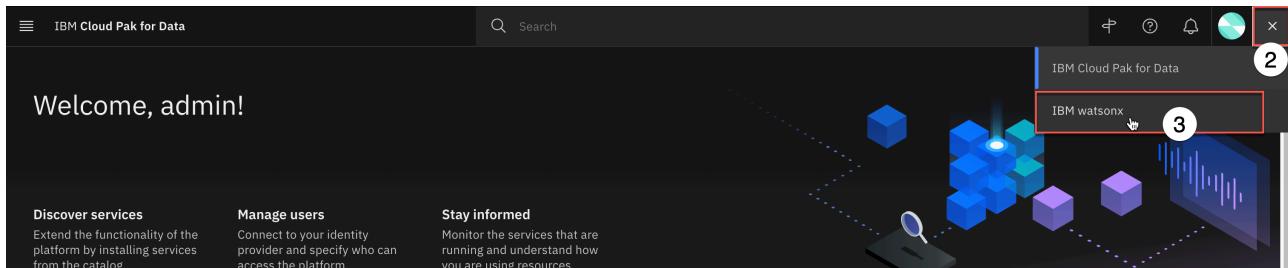


## 2. Switch contexts

The [IBM watsonx](#) context offers an improved user interface and better integration for AI governance than the Cloud Pak for Data context, and offers expanded functionality such as monitoring for detached prompt templates. Some operations, such as creating a database, currently require using the [Cloud Pak for Data context](#). However, for the remainder of the lab, you will use the [IBM watsonx](#) context.

1. Log into the Cloud Pak for Data home page using the credentials from your reservation.
2. Click on the **grid icon** in the upper right to open the context menu.

3. Click on the **IBM watsonx** menu item to change the context. A **Welcome to watsonx** popup window may open.



4. Close the popup window, or click the **Take a tour** button if you wish.

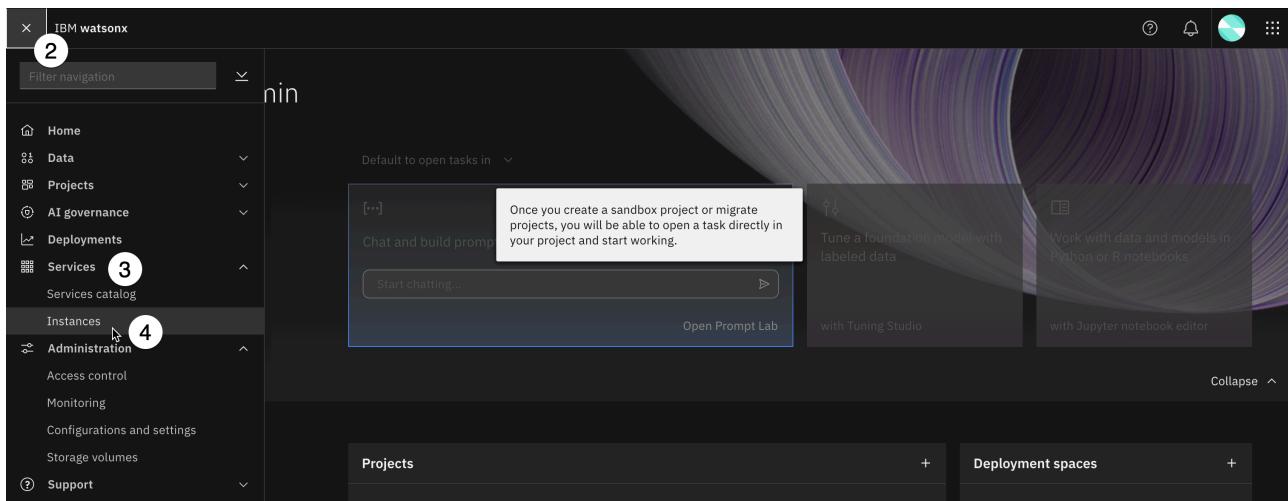
Your screen should now show IBM watsonx branding. This screen will be referred to throughout the lab as the watsonx home screen.

**Note:** You may need to switch to the watsonx context if you log out and then log back into your environment.

### 3. Launch the governance console

In this section, you will launch the OpenPages service.

1. If necessary, return to the watsonx home page by clicking the **IBM watsonx** link in the upper left.
2. Click on the **hamburger menu** in the upper left.
3. Click on the **Services** item from the menu to expand it.
4. Click on **Instances** to open the **Instances** screen.



5. Locate the instance of OpenPages in the table and click on the link in the **Name** column to open the instance details screen.

The screenshot shows the 'Instances' section of the IBM Watsonx interface. At the top, there are filters for 'Type', 'Status', 'Data plane', and 'Physical location'. A search bar says 'Find instances'. On the right, there are buttons for 'New instance' and a '+' sign. Below the header is a table with columns: Name, Type, Created by, vCPU requests, Memory requests (GiB), Data plane, Physical location, Status, and Created on. The table contains three rows:

Name	Type	Created by	vCPU requests	Memory requests (GiB)	Data plane	Physical location	Status	Created on
cpd-database Service instance for db2oltp-17...	db2oltp	admin	2.10	4.25 Gi	—	—	<span>Green</span>	Oct 11, 2024
openscale-defaultinstance IBM Watson OpenScale	aios	admin	0.00	0.00 Gi	—	—	<span>Green</span>	Oct 8, 2024
openpagesinstance-cr OpenPages Instance	openpages	admin	4.45	12.40 Gi	—	—	<span>Green</span>	Oct 8, 2024

6. Scroll down to the [Access information](#) section, and click the [launch icon](#) to launch the service.

The screenshot shows the configuration page for the 'openpagesinstance-cr' service. It has sections for 'Status' (Running), 'Database configuration', 'Access information', and 'Size'. In the 'Access information' section, there is a 'URL' field containing the value 'https://cpd-cpd.apps.ocp-110000b3qc-p09m.cloud.techzone.ibm.com/openpages-openpagesinstance-cr/'. Below the URL is a 'Launch OpenPages' button, which is highlighted with a black box and a circled number 6. Other fields in this section include 'Database type' (Internal database), 'Use dedicated nodes' (False), and 'Node label' (Node label). The 'Size' section shows 'Data storage class' as 'ocs-storagecluster-ceph-rbd'.

The watsonx governance console (OpenPages service) launches.

## 4. Change user profiles

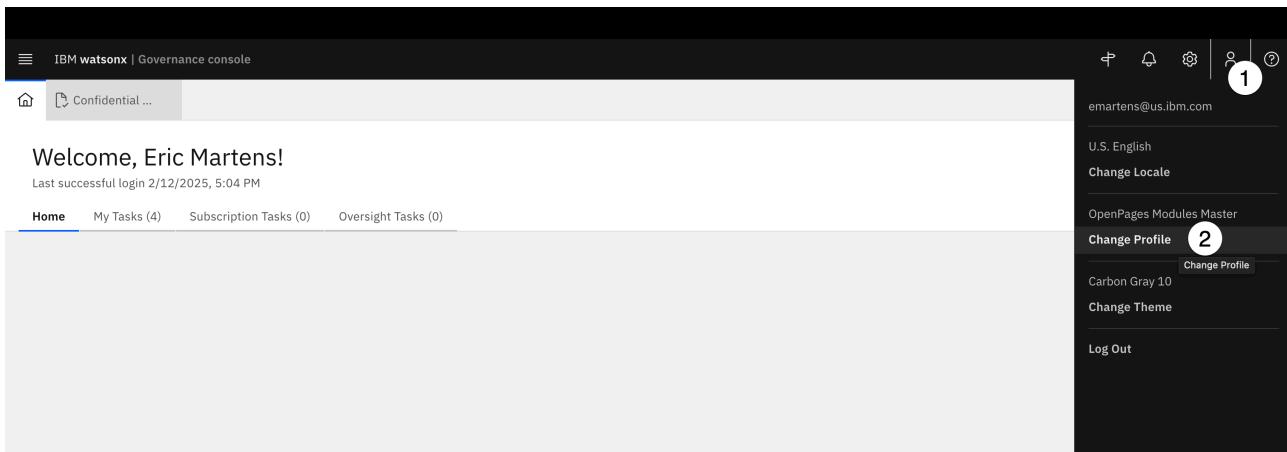
Access and views in the governance console is managed via user profiles. As with anything in the console, it can be customized to fit an organization's needs. Each user can have as many profiles assigned as needed, and they can switch between those profiles depending on the task they need to accomplish.

A user's profile primarily determines the information present in the various views, which allows administrators to configure the tool so that each user sees exactly the information they need, without being overwhelmed with unnecessary options.

For this lab, you have been assigned all the available pre-configured roles for watsonx.governance. Again, in a real-world situation, a user would likely only have one or two profiles, but that level of user management is beyond the scope of this lab.

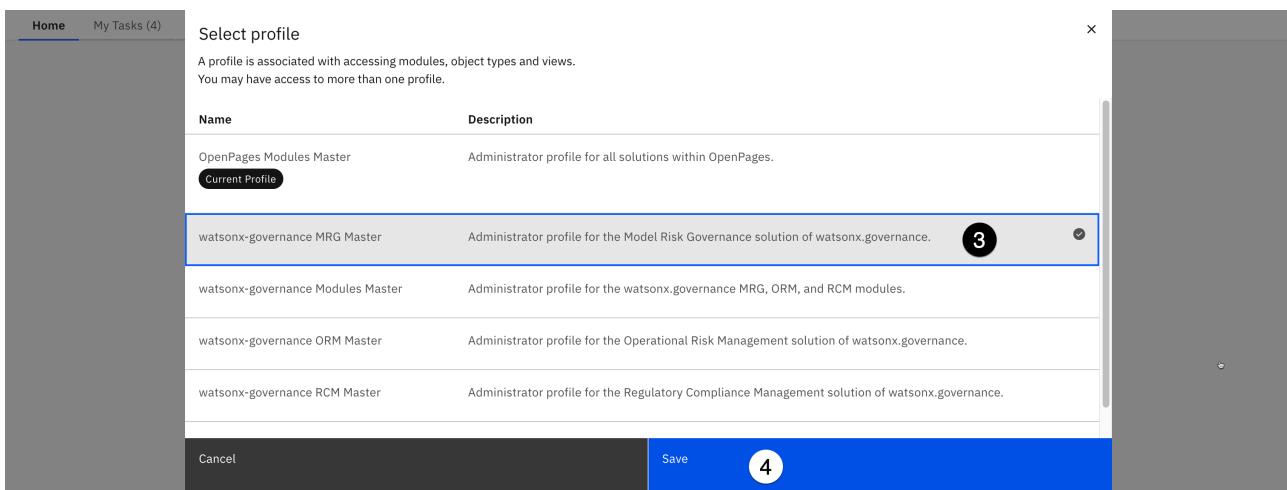
To continue, you will need to change your role to that of watsonx.governance MRG (Model Risk Governance) Master.

1. Click on the [avatar](#) icon in the upper left of the screen. The user menu opens.
2. The current profile likely shows as [OpenPages Modules Master](#). Click on the [Change Profile](#) menu item. The [Select profile](#) dialog opens.



3. Click on the [watsonx-governance MRG Master](#) profile from the list to select it.

4. Click on the [Save](#) button to save your choice.



Each time you sign in, you should now sign in as the watsonx-governance MRG Master profile.

## 5. Create a model use case

To ensure that model use cases are tracked across the entire solution, they should be created using the watsonx governance console. During the configuration of the environment you are using, the administrator turned on integration between the governance console (OpenPages) and watsonx, so any actions related to model use cases should now redirect you to the governance console interface.

Only the models that you add to use cases are tracked with AI Factsheets. You can control which models to track for an organization without tracking samples and other models that are not significant to the organization.

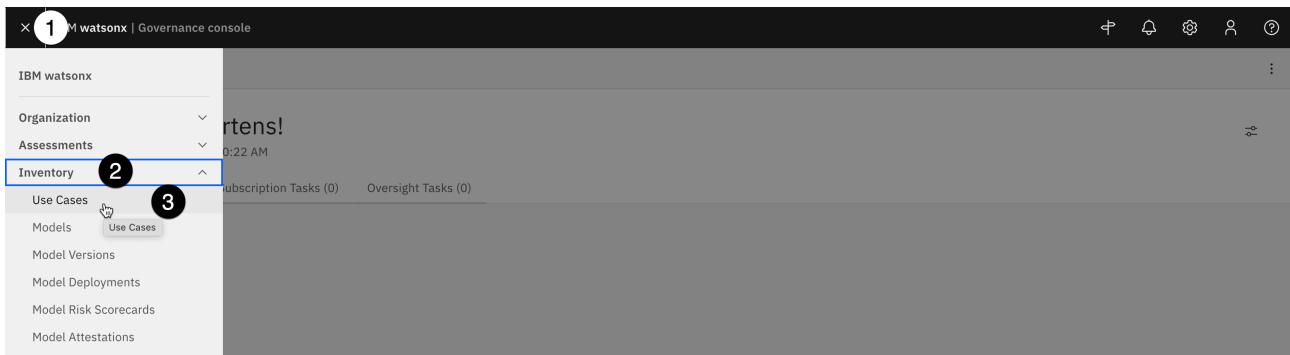
In a real-world scenario, this action would be performed by an organizational stakeholder who would like to request the development and implementation of a model; in this case, the manager of the human resources department, who is unable to keep up with the volume of resumes submitted for employment opportunities and would like help from an AI solution.

### Choosing a use case

In this section, of the lab, you will create a use case based on a business need. As part of the process, you will then fill out a questionnaire to identify potential risks associated with that use case. Feel free to create a use case from previous experience you had with a client. Take a moment to think of a business problem

that a client would use AI or machine learning to solve. You may also use one of the [sample use cases](#) provided for you. When you have identified a use case, proceed with the lab.

1. Click on the [hamburger menu](#) in the upper left.
2. Click on the [Inventory](#) menu item to expand it.
3. Click on the [Use Cases](#) menu item. The [Use Cases](#) tab opens. Note that this tab will be empty, but will populate as others doing the lab create use cases in this environment.



4. Click the blue [New](#) button. The [New Use Case](#) tab opens.

A screenshot of the 'Use Cases (33)' list view. The page has a header with a search bar and a 'View Name' dropdown. Below is a table with columns: Name, Purpose, Description, Owner, Status, Risk Level, and Tags. Three rows of data are visible:

Name	Purpose	Description	Owner	Status	Risk Level	Tags
Agency Based LGD Estimation	High Oaks Bank > North America > Corporate Banking	Uses internal and external recovery data, adjusted for macro-economic impact. Uses statistical regression	Bob Eldridge	Approved for Development	Low	
Banking book HTM corporate bond - income	High Oaks Bank > Europe > Corporate Banking	ALM based income forecast for the HTM portfolio, initially for the CCAR 2013 stress-test. Vendor solution using conditional scenarios and core ALM system.	Bob Eldridge	Approved for Development	Medium	

A blue 'New' button with a white plus sign is located in the top right corner of the table header. A small black circle with the number '4' is placed over this button.

Note that the [Model Use Case creation](#) information panel on the right of the screen offers helpful information about model use cases, as well as a list of required fields. Clicking on any of the fields in that panel will scroll the screen directly to that portion of the form, helping you quickly rectify any items needing attention.

5. In the [General](#) section of the form, enter the name of your use case in the [Name](#) field. Note that when you enter a value in the field, the progress bar in the [Model Use Case creation](#) information panel updates. For this lab, it may be helpful to include your name or email address in the use case name, as some of your colleagues may be creating similar use cases.
6. Click the [Owner](#) field and enter the name of the user you signed in with into this field.
7. Enter a description in the [Description](#) field.

\*Modified Required\*

### General

*Name *	*Owner *
Resume summarization emartens	emartens@us.ibm.com

Purpose

*Description *
Summarize resumes from job applicants

**Model Use Case creation**

A **model use case** is meant to track and capture information about a collection of models that will be built to serve a particular purpose. A model use case should be created whenever there is a

more

1 item requires attention.

All Key Items (5) ▾

 Name \* Owner \*

8. All model use cases are owned by business entities, representing the part of the organization responsible for requesting the use case. In the **Business Entities** section of the form, click the **Add** button. The **Add** window opens with a list of business entities defined for the organization.

Name	Description	Entity Type	Tags
No results			

9. The list of business entities includes a variety of corporate departments such as **Human Resources**, **Legal**, and **Procurement**. Choose one that best fits your use case from the list and click on it to select it.

10. Click **Done** to add the business entity to the use case. The **Add** window closes.

Foundation Models	
Fulfillment	The fulfillment department for GloboCorp.
GlobalCorp	A global institution with operations across every continent, offering a broad range of services.
<input checked="" type="checkbox"/> Human Resources	The human resources department for GlobalCorp.
Legal	The legal department for GloboCorp.
Library	Library
MRG	Library > MRG
Procurement	The procurement department for GloboCorp.
PCM	PCM

Cancel Done

11. Click the **Save** button in the upper right to save the use case.

To progress the use case through the workflow, you will now need to perform the actions specified in the **Action** items in the workflow.

6. Progress the use case to the next phase

The use case request has progressed to the data gathering stage of the workflow, and has been assigned as an action for the appropriate owner. Recall that owners of each stage of the workflow can be configured, and alerts assigned.

1. Click on the [Home](#) icon in the upper left to return to the user's home tab.

Name	Status	Stage	Due Date	Description
Use Case Data Gathering (Data gathering)	Proposed	Data gathering	5/29/2024	(Data gathering)

2. Note that the [My Tasks](#) tab now shows a new entry. Click on the tab to open it.

Name	Due Date
Use Case Data Gathering (Data gathering)	5/29/2024

The [My Tasks](#) tab shows a list of all the current tasks assigned to the user. It can be filtered by a variety of fields. At the moment, it only contains a single task, showing that the use case request is in the data gathering stage and is in need of action, along with the stage due date.

3. Click on the link for the use case in the table to return to the use case request tab.

Name	Type	Workflow Name	Stage (Status)	Criticality	Stage Due Date
<a href="#">Resume summarization</a>	Use Case	Use Case Request	Use Case Data Gathering (Data gathering)	Medium	5/29/2024

The use case request form shows a standard set of information. However, like all views and forms in the governance console, it can be customized to display data an organization would like to track as part of this process. Information could include things like billing codes, additional documentation or justification, or more. In this case, you will only edit required fields specified in the information panel on the right before progressing to the next stage of the workflow.

**Risk Level** represents the risk to the organization should issues arise with the models used to address the requirements laid out by the use case. A full risk assessment is beyond the scope of this lab; however, because hiring and employment violations can lead to expensive litigation damage to an organization's

reputation, this use case in this example (resume summarization) will be marked as high risk. You may choose the risk level most appropriate to your use case.

4. In the **Risk** section of the form, click on the pencil icon next to the **Risk Level** field to edit it.

The screenshot shows the 'Risks' section of a use case form. At the top, there's a search bar and a 'Copy from Library' button. Below is a table with columns: Name, Description, Inherent Risk Rating, Residual Risk Rating, Status, and Tags. A message says 'No results'. On the right, there's an information panel for the stage 'Use Case Data Gathering (Data gathering)' with fields for Stage, Due Date, and Tags. A note at the bottom says 'Please capture all relevant information to this AI use case proposal and then submit using the Action button'.

5. Select the appropriate risk level from the dropdown.

6. Scroll to the **General** section and click on the pencil icon next to the **Purpose** field to edit it.

7. Fill out a business purpose for your use case.

8. Click the **Save** button in the upper right to save your changes.

9. Once the changes have been saved, click on the **Action** button in the upper right to open the **Actions** menu.

10. Click on the **Submit for initial approval** action. Note that the text of this action is defined by the **Name** field given to the action connecting the **Use Case Data Gathering** stage to the **Initial Approval** stage in the workflow. The **Submit for initial approval** confirmation dialog opens.

The screenshot shows the 'General' section of a use case form. It includes fields for Name (Resume summarization emartens), Status (Proposed), and Description (Summarize resumes from job applicants). On the right, there's an information panel for the stage 'Use Case Data Gathering (Data gathering)' with fields for Stage, Due Date, and Tags. A circled number 9 is over the 'Action' button, and a circled number 10 is over the 'Submit for initial approval' option in the dropdown menu.

12. Click the black **Continue** button to confirm your action, but keep the use case tab open.

**Note:** If you receive a **Network error** message, your change may have been recorded, but network issues may have prevented the screen from refreshing. Try submitting again; if the error persists, click the **Refresh workflow info** button to the right of the **Stage** field in the information panel on the right. The **Stage** should progress to **Initial Approval (Awaiting use case approval)**.

The screenshot shows the IBM Watsonx Governance console interface. At the top, there's a navigation bar with 'IBM Watsonx | Governance console'. Below it, a header bar has tabs for 'Use Cases' (which is selected) and 'Resume sum...'. A search bar is also present. The main content area shows a use case card for 'Resume summarization'. The card includes fields for 'Name' (Resume summarization), 'Use Case Type' (AI), 'Status' (Awaiting Use Case Approval), 'Owner' (complianceofficer), and 'Purpose'. To the right of the card is an 'Actions' button. A tooltip for 'Refresh workflow info' is displayed over the actions menu, with a red box highlighting the 'C' icon in the tooltip.

- When the action completes, note that the **Stage** field in the information panel on the right has updated once again to **Initial Approval**.

In order to progress the use case to the next stage (**Stakeholder Review**) the action in the workflow requires the risk identification questionnaire to be filled out. As that questionnaire has yet to be completed, clicking on the **Actions** menu for the use case only shows two available actions: rejecting the use case (moving it to the **Rejected** stage, or returning it to the owner (moving it back to the **Use Case Data Gathering** stage). In order to continue forward, the questionnaire must be filled out.

## 7. Identify use case risks

In this section, you will fill out the default risk assessment questionnaire included in the governance console. This questionnaire, which can be modified to suit an organization's needs and regulatory requirements, has been configured to automatically associate relevant risks from IBM's AI risk atlas based on answers to the questions.

The **AI risk atlas** is an open source tool to help clients understand some of the risks of working with generative AI, foundation models, and machine learning models.

**Note:** If you receive frequent error messages stating that *The requested operation could not be completed*, you are likely encountering an issue with persistent session information in your browser. A browser cache clear may fix this issue, but the best way to avoid these errors is to use your browser's private/incognito mode when signed in as the created user.

- Click on the **Home** tab.
- Click on the **My Tasks** tab from the home screen to reopen the view of assigned tasks.
- Click on the **AI Risk Identification ...** task from the task list that corresponds to the use case you created. The **Risk Identification** questionnaire assessment for the use case opens.

Welcome, Attendee 29!

...

Home My Tasks (3) Subscription Tasks (0) Oversight Tasks (0)

2

My Tasks

3

Filter By: Criticality Workflow Name Stage Type Stack By: Type View By: Week

wk of 9/15 Now wk of 9/22 No Due Date Upcoming

Name	Type	Workflow Name	Stage (Status)	Criticality	Stage Due Date
Resume summarization - attendee29	Use Case	Use Case Request	Initial Approval (Awaiting use case approval)	Medium	9/26/2025
AI Risk Identification (Resume summarization - attendee29)	Questionnaire Assessment	Questionnaire Assessment Workflow	Information Gathering	Medium	
Applicability Assessment (Resume summarization - attendee29)	Questionnaire Assessment	Questionnaire Assessment Workflow	Information Gathering	Medium	

4. Fill out the questionnaire for your use case. Your answers should reflect real risks posed by using AI models to address this business issue. For example, when summarizing resumes, input data will be provided by humans, and may contain some sensitive information. Note that the idea of this questionnaire is to identify potential risks in the model use case. Feel free to use short, generic answers when filling out text entry portions of the form such as specifying things like bias mitigation strategies or describing content.

**Note** The purpose of this questionnaire is NOT to provide an out-of-the-box solution to identifying client risks. Instead, it is to provide a foundation on which clients or consultants can build their own questionnaires, customized to a particular client. As you progress through the questionnaire, it may be useful to consider other questions that might be relevant for some of your clients. They can easily modify existing forms, or build new ones from scratch using the editor. Also note how additional questions can be added to the form based on answers to other questions.

Some of the answers require comments to be added; when adding those comments, you must click on the blue **Submit** button to the right of the text entry field for the comment to be entered and the question marked as completed.

View all questions

Questions completed  
50/51

Sections

Risk Identification

- Use Case
- Use
- Context

Additional Risks

- Data Risks
- Output Risks

Additional Risks 0/1

1.4.1. Human Verification

Will a human have to verify the model's output before it is used?  
Please describe. \*

Comment  Attachment  Activity

Description here

Note the progress panel on the left side of the screen will show any required questions that have not been answered, and you can use it to jump between sections of the questionnaire. Your progress will also be automatically saved as you answer questions.

5. When you have finished filling out the survey, click the **Action** button in the upper right. The **Actions** menu opens.
6. Click the **Submit and Close** button. A confirmation dialog opens.

7. Click **Submit** to submit the risk identification questionnaire.

8. Based on the questionnaire answers, the governance console now calculates and assigns certain risks to the use case. You can view these by clicking on the **Home** tab.

9. Click on the **My Tasks** tab of the **Home** tab. The use case appears in the list of tasks, with the stage set to **Initial Approval**.

10. Click on the use case from the task list. The use case opens in a new tab.

11. Scroll down to the **Risk** section of the page.

**Note:** If you do not see any risks associated with your use case, you may need to refresh the browser window to populate the table.

12. Scroll down to the **Risk Status** and **Residual Risk Rating** graphs. Based on your questionnaire answers, your charts may look different from the screenshot. Click on the **Risk Status** graph. The **Risks** tab opens.

13. Examine the table of risks. Note that each has a description, and a reference URL for more information. These risks have been populated from the [AI risk atlas](#). They also have a **Status** of **Awaiting assessment**, indicating that a risk assessor must decide if they are relevant to the use case or not.

Next, you will assess the individual risks from the role of a risk manager.

## 8. Assess individual risks

The questionnaires you completed have been constructed to automatically add various risks to the use case, based on the answers provided. Clients looking to use a similar process can use this questionnaire template as a model for creating their own, customized to their individual use cases.

In this section, you will assess individual risks. For the sake of time, you will only perform a single in-depth assessment, to see how this is handled in the governance console.

1. The **Risks** tab for the use case should be open and active on your screen. Choose one of the risks on the table and click on it to open it.

Risks (17) Related to Resume summarization emartens Status : Awaiting Assessment

Name	Description	Owner	Status	Risk Category	Inherent Risk Rating	Residual Risk Rating	Tags
<a href="#">Confidential Information in data (MOD_0000000_RIS_0000009)</a> GlobalCorp > Human Resources	Models might be trained or fine-tuned using confidential data or the company's intellectual property, which could result in unwanted disclosure of that information.	System Administrator	Awaiting Assessment		Not Determined	Not Determined	
<a href="#">Confidential data in prompt (MOD_0000000_RIS_0000018)</a> GlobalCorp > Human Resources	Inclusion of confidential data as part of a generative model's prompt, either through the system prompt design or through the inclusion of end user input, might later	System Administrator	Awaiting Assessment		Not Determined	Not Determined	
<a href="#">Dangerous use (MOD_0000000_RIS_0000037)</a> GlobalCorp > Human Resources	The possibility that a model could be misused for dangerous purposes such as creating plans to develop weapons, malware, or causing harm to others is the risk of	System Administrator	Awaiting Assessment		Not Determined	Not Determined	

2. Scroll down to the **Related Content** section of the page, and note that the risk can be associated with mitigating controls, processes, or other issues. Take a moment to inspect some of the other sections on the page, including **Internal Audit Risk Rating**, and note how risks can be customized based on the threat they pose to a client's business.

3. Click on the **Action** button in the upper right to open the actions menu.

4. Click on the **Start model risk assessment** button to begin assessing the risk. A confirmation dialog opens.

Risk  
Confidential Information i...    \*    \*

Inherent Risk Rating Not Determined    Residual Risk Rating Not Determined    Owner System Administrator

Action Start Model Risk Assessment

Task   Activity   Admin

\* Modified Required \*

General

Name Confidential Information in data (MOD\_0000000\_RIS\_0000009)  
Description Models might be trained or fine-tuned using confidential data or the company's intellectual property, which could result in unwanted disclosure of that information.  
Owner System Administrator

Status   Assessment Method

Tags No tags have been added yet.

Risk awaiting evaluation Assess the risk by reviewing the information provided / Controls / User Events / KPIs / Processes

5. Click the **Continue** button. The risk assessment form opens. If you look at the task list on your home tab, you will also see the assessment there.

Note that you have been marked as the **Owner** of the risk assessment. In the information panels on the right side of the screen, you will see a **Confirm Assignment** panel, which asks you to confirm that this assignment is correct. Next, you will need to either mark the risk as ready for assessment, or set it as not applicable to the use case.

7. Click on the **Actions** button. The **Actions** menu opens.

8. Click on the **Ready for Assessment** menu item. A confirmation dialog opens.

Risk  
Confidential Information i... ☆ ^

Inherent Risk Rating  
Not Determined

Residual Risk Rating  
Not Determined

Owner  
emartens@us.ibm.com

**Actions** 7

Ready for Assessment 8

Risk Not Applicable

**Task** Activity Admin

\*Modified Required\*

**General** ⓘ

Name Confidential Information in data (MOD\_0000000\_RIS\_0000009)  
Owner emartens@us.ibm.com

Description \*  
Models might be trained or fine-tuned using confidential data or the company's intellectual property, which could result in unwanted disclosure of that information.

Domain Assessment Method

9. Click on the **Continue** button in the confirmation dialog. The risk assessment form changes to show the assessment values.

10. Scroll down to the **Risk Assessment** portion of the page and click on the **information icon** next to the session header to open the **Field Guidance** window. Take a moment to read the descriptions of what each field represents.

Exposing Personal Information (MOD\_0000000... ☆ ^)

Action

**Task** Activity Admin

\*Modified Required\*

**10** Risk Assessment ⓘ Field guidance

Inherent Impact Not Determined

Mitigation Strategy

Residual Impact Not Determined

Inherent Likelihood Not Determined

Residual Likelihood Not Determined

Inherent Risk Rating Not Determined

Residual Risk Rating Not Determined

Tags

No tags have been added yet.

**Perform Risk Assessment** ⓘ

Perform Risk assessment by updating the following data:

11. Close the **Field Guidance** window by clicking the **X** button in the upper right corner of the popup.

12. Click on the **edit icon** for each field and assign a rating.

Risk Assessment ⓘ

\* Inherent Impact Medium

\* Residual Impact Medium

\* Inherent Likelihood Low

Residual Likelihood Not Determined

\* Inherent Risk Rating Low

\* Residual Risk Rating Low

**Monitoring & Mitigation**

Controls

Search

Name	Description	Control Owner	Control Type	Operating Effectiveness	Status	Tags
------	-------------	---------------	--------------	-------------------------	--------	------

No tags have been added yet.

**Perform Risk Assessment** ⓘ

Perform Risk assessment by updating the following data:

- Inherent Impact and Likelihood

Select an action to validate

All Key Items (6) ▾

Inherent Impact

Inherent Likelihood

Next, you will need to provide a mitigation strategy for the risk.

13. Locate the **Mitigation Strategy** field in the **Risk Assessment** section. Hover your mouse over it and click on the **pencil icon** to edit it.

Risk Assessment

Inherent Impact: Medium

Inherent Likelihood: Low

Inherent Risk Rating: Low

Mitigation Strategy 13

Residual Impact: Medium

Residual Likelihood: High

Residual Risk Rating: Medium

Monitoring & Mitigation

No tags have been added yet.

Perform Risk Assessment

Perform Risk assessment by updating the following data:

- Inherent Impact and Likelihood

Select an action to validate

All Key Items (7)

14. Enter a strategy in the text field for dealing with the risk. For example, for a *Confidential information in data* risk for a resume summarization use case, a strategy could involve masking or obfuscating sensitive info using watsonx.data.

15. Click the **Save** button to save your changes.

Confidential Information in data (MOD\_000000...)

Cancel Save 15

Task Activity Admin

\*Modified Required\*

Risk Assessment

\* Inherent Impact: Medium

\* Inherent Likelihood: Low

\* Inherent Risk Rating: Low

\* Residual Impact: Medium

\* Residual Likelihood: High

\* Residual Risk Rating: Medium

Monitoring & Mitigation

Perform Risk Assessment

Perform Risk assessment by updating the following data:

16. Click on the **Action** button in the upper right to open the actions menu.

17. Click on the **Assessment Complete** button to finish the risk assessment.

IBM Watsonx | Governance console

Confidential Information in data (MOD\_000000...)

Action 16

Assessment Complete 17

Task Activity Admin

\*Modified Required\*

Risk Assessment

Inherent Impact: Medium

Inherent Likelihood: Low

Inherent Risk Rating: Low

Residual Impact

Residual Likelihood

Residual Risk Rating

Tags

No tags have been added yet.

Perform Risk Assessment

Perform Risk assessment by updating the following data:

18. When asked to confirm your choice, click on the **Continue and close tab** button.

19. Return to the use case view, either by clicking on the tab or locating it from the **My tasks** section of your **Home** tab.

20. Scroll down to the **Risks** section, and note that the **Inherent Risk Rating**, **Residual Risk Rating**, and **Status** have been updated in the table.

You may repeat this process for as many of the risks as you wish before proceeding. For the sake of brevity, the next steps show you how to change the status of multiple risks at once.

## 9. Assess multiple risks

The defined workflow for use case requests dictates that the use case cannot pass to the next phase until all the associated risks have been assessed.

- Click on the **Launch Grid page** button at the top of the **Risks** table. The grid page opens.

Name	Description	Inherent Risk Rating	Residual Risk Rating	Status	Tags
Data bias (MOD_0000000_RIS_0000001)	Historical, representational, and societal biases present in the data used to train and fine tune the model can adversely affect the model's performance.	Not Determined	Not Determined	Awaiting Assessment	
Data poisoning (MOD_0000000_RIS_0000002)	Data poisoning is a type of adversarial attack where an adversary or malicious insider injects intentionally corrupted, biased data into the system.	Not Determined	Not Determined	Awaiting Assessment	

- Check the box to the left of all the risks still marked with the **Awaiting Assessment** status. Note that you may need to scroll the window to check them all.

- Click the **Bulk Update** button at the top of the table. The **Bulk Update** panel opens.

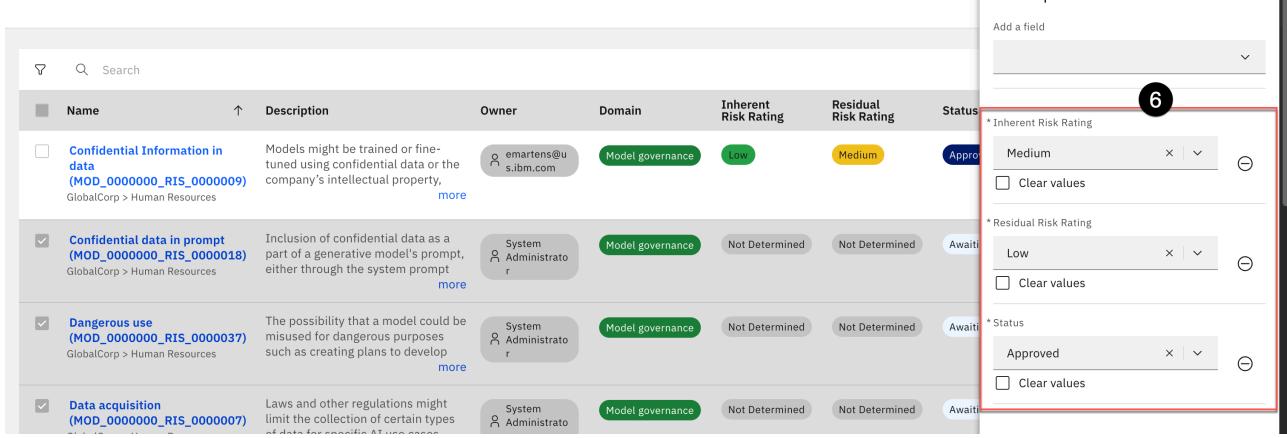
Name	Description	Owner	Domain	Inherent Risk Rating	Residual Risk Rating	Status	Reference URL	Tags
Confidential Information in data (MOD_0000000_RIS_0000009)	Models might be trained or fine-tuned using confidential data or the company's intellectual property.	emartens@us.ibm.com	Model governance	Low	Medium	Approved	Risk Atlas (Confidential Information in data)	
Confidential data in prompt (MOD_0000000_RIS_0000018)	Inclusion of confidential data as a part of a generative model's prompt, either through the system prompt.	System Administrator	Model governance	Not Determined	Not Determined	Awaiting Assessment	Risk Atlas (Confidential data in prompt)	
Dangerous use (MOD_0000000_RIS_0000037)	The possibility that a model could be misused for dangerous purposes such as creating plans to develop weapons.	System Administrator	Model governance	Not Determined	Not Determined	Awaiting Assessment	Risk Atlas (Dangerous use)	
Data acquisition (MOD_0000000_RIS_0000007)	Laws and other regulations might limit the collection of certain types of data for specific AI use cases.	System Administrator	Model governance	Not Determined	Not Determined	Awaiting Assessment	Risk Atlas (Data acquisition)	
Data bias (MOD_0000000_RIS_0000001)	Historical, representational, and societal biases present in the data.	emartens@us.ibm.com	Model governance	Not Determined	Not Determined	Awaiting Assessment	Risk Atlas (Data bias)	

- Click on the **Add a field** dropdown and select the **Inherent Risk Rating** item from the list. A dropdown for **Inherent Risk Rating** appears in the panel.

Name	Description	Owner	Domain	Inherent Risk Rating	Residual Risk Rating	Status
Confidential Information in data (MOD_0000000_RIS_0000009)	Models might be trained or fine-tuned using confidential data or the company's intellectual property.	emartens@us.ibm.com	Model governance	Low	Medium	Approved
Confidential data in prompt (MOD_0000000_RIS_0000018)	Inclusion of confidential data as a part of a generative model's prompt, either through the system prompt.	System Administrator	Model governance	Not Determined	Not Determined	Awaiting Assessment
Dangerous use (MOD_0000000_RIS_0000037)	The possibility that a model could be misused for dangerous purposes such as creating plans to develop weapons.	System Administrator	Model governance	Not Determined	Not Determined	Awaiting Assessment

- Repeat the previous step to add **Residual Risk Rating** and **Status** to the panel.

- Click on the dropdowns and select risk ratings and a status. Note that to progress the use case, you must select either **Approved** or **Not Applicable** in the **Status** dropdown.

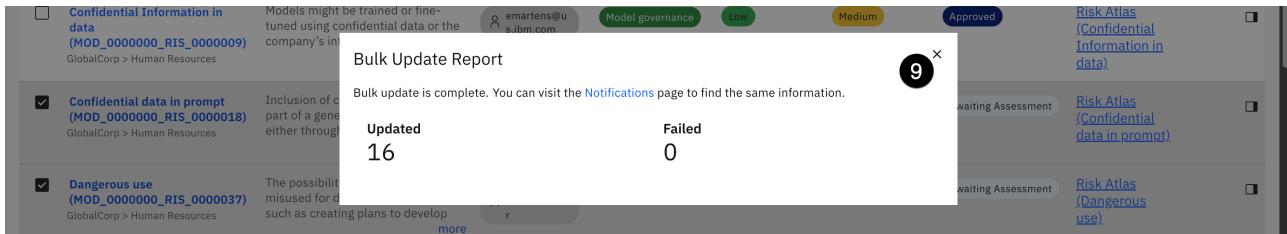


Name	Description	Owner	Domain	Inherent Risk Rating	Residual Risk Rating	Status
<input type="checkbox"/> Confidential Information in data (MOD_0000000_RIS_000009)	Models might be trained or fine-tuned using confidential data or the company's intellectual property.	emartens@us.ibm.com	Model governance	Low	Medium	Approved
<input checked="" type="checkbox"/> Confidential data in prompt (MOD_0000000_RIS_000018)	Inclusion of confidential data as a part of a generative model's prompt, either through the system prompt	System Administrator	Model governance	Not Determined	Not Determined	Awaiting Assessment
<input checked="" type="checkbox"/> Dangerous use (MOD_0000000_RIS_000037)	The possibility that a model could be misused for dangerous purposes such as creating plans to develop	System Administrator	Model governance	Not Determined	Not Determined	Awaiting Assessment
<input checked="" type="checkbox"/> Data acquisition (MOD_0000000_RIS_000007)	Laws and other regulations might limit the collection of certain types of data for certain AI use cases.	System Administrator	Model governance	Not Determined	Not Determined	Awaiting Assessment

7. Click the **Update** button at the bottom of the panel.

8. When asked to confirm your bulk update, click the **Confirm** button. The update will run, and may take a few minutes to complete depending on how many risks were updated.

9. When the update completes, click the **X** button to close the **Bulk Update Report** popup. The **Risks** table will refresh, showing the new values.



The risk assessments are now complete. You can progress the use case to the next phase.

## 10. Assess applicability for the use case

In this step, fill out the applicability assessment questionnaire. As with the previous questionnaire, this form is provided by IBM, but can be fully customized by clients to fit their own needs. It is modeled after some of the requirements set forth in legislation around AI in the European Union.

1. Click on the **Home** tab to switch to it. The tab should still be showing the **My Tasks** view.
2. Locate and click on the **Applicability Assessment** link from the tasks table. The **Applicability Assessment Questionnaire** opens in a new tab.

The screenshot shows the IBM Watsonx Governance console dashboard. At the top, there's a header with the title 'IBM Watsonx | Governance console'. Below the header, there are navigation links for 'Use Cases' and 'Resume sum...'. A circular callout '1' is positioned above the 'Use Cases' link.

The main area is titled 'Welcome, Eric Martens!' and shows a message 'Last successful login 10/14/2024, 7:18 PM'. Below this, there are tabs for 'Dashboard', 'My Tasks (2)', 'Subscription Tasks (0)', and 'Oversight Tasks (0)'. The 'My Tasks (2)' tab is selected.

The 'My Tasks' section displays a chart titled '2' showing task counts for 'Questionnaire Assessment' (blue bar) and 'Use Case' (purple bar). The chart indicates one task is 'Now' and one has 'No Due Date'. Below the chart is a table with columns: Name, Type, Workflow Name, Stage (Status), Criticality, and Stage Due Date. The first row is for 'Resume summarization' (Type: Use Case, Status: Initial Approval (Awaiting use case approval)). The second row is for 'Applicability Assessment (Resume summarization)' (Type: Questionnaire Assessment, Status: Applicability Assessment). A circular callout '2' is positioned above the second row.

Take a moment to skim the text for each of the questions, and understand how potentially clients might answer them and how it affects their risk profile for adopting AI tools.

After the third question, the questionnaire will add extra questions based on your answers. Continue to fill out the questions for the resume summarization use case, taking into account whether the model will perform classification, facial or image recognition, biometric data, individual risk assessments, or other potentially harmful acts.

When you have completed the questions, the [Category Assessment](#) section of the form will populate.

### 3. Click on the [Category Assessment](#) section to view it.

The screenshot shows the 'Questionnaire Assessment' page. At the top, there are tabs for 'Task', 'Activity', 'Admin', and 'Questionnaire'. The 'Questionnaire' tab is selected. There are sections for 'View all questions', 'Questions completed', and a date '9/21'. On the left, there's a sidebar with 'Sections' expanded, showing 'Applicability Assessment' (selected), 'Scope and Prohibited AI Systems', and 'Category Assessment' (with a circular callout '3').

The main content area shows a question: 'Does the AI System include the placing on the market, putting into service or use of an AI system that: exploits any of the vulnerabilities of a person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective, or the effect, of materially distorting the behaviour of that person or a person belonging to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm? \*'. Below the question are 'Yes' and 'No' radio buttons, with 'No' selected. There are also 'Clear' and 'Action' buttons at the bottom right.

### 4. Continue to fill out the questions in the questionnaire for your use case.

### 5. When you have finished answering all the questions, click on the [Action](#) button to open the actions menu.

### 6. Click on the [Applicability assessment complete](#) menu item.

Questionnaire Assessment  
Applicability Assessment (Resume summariza... ☆ ⓘ ^

Assignee Stage Name  
Applicability Assessment Save draft Action

Task Activity Admin Questionnaire

View all questions Questions completed 23/23

You are: \*

- a provider of an AI system that interacts directly with natural persons (e.g. chatbots)
- a provider of an AI system, including general-purpose AI systems, generating synthetic audio, image, video or text content
- a deployer of an emotion recognition system or a biometric categorisation system
- a deployer of an AI system that generates or manipulates image, audio or video content constituting a deep fake
- None of the above

Comment Attachment Activity

2.1.14. Contact Compliance Department

7. Click on the **Submit** button to confirm your choice. Your action may take a few moments to save, and the screen may not update. However, you may proceed with the lab.

You have now completed two assessments regarding the model use case, which have been used to both automatically identify possible risks associated with using AI and helped insure regulatory compliance. Next, you will approve the use case for development.

## 11. Approve the use case for development

Now that the risks have been identified and assessed, the use case can be approved for the next stage of the lifecycle.

1. Return to the use case view, either by clicking on the tab or locating it from the **My tasks** section of your **Home** tab.
2. The use case is now ready to be progressed to the next stage of the workflow. Click on the **Actions** button in the upper right. The **Actions** menu opens.
3. Click on the **Submit for stakeholder review** menu option. A confirmation dialog opens.

IBM Watson | Governance console

Use Case Resume summarization ☆ ^

Status: Awaiting Use Case Approval Risk Level: High

Task Activity Admin

\* Modified Required \*

General ⓘ

Name \* Resume summarization Use Case Type: AI Status: Awaiting Use Case Approval

Description Owner Purpose

Stage: Initial Approval (Awaiting use case approval) Due Date: 10/20/2024 Tags:

Actions

- Reject use case
- Submit for stakeholder review**
- Return to owner

4. Click on the **Continue** button to confirm your choice. The use case progresses to the **Stakeholder Review** stage.

At this point in the process, the model risk department would review the use case, including the answers provided in the risk identification questionnaire. Other stakeholders would also review the use case to ensure that all requirements were being met. Note that a link to the questionnaire have been provided in the **Risk** section of the page, for easy access. Also note that the bar chart of the risks has been updated to reflect their assessments.

5. Click on the **Actions** button once more. The **Actions** menu opens.
6. Click on the **Approve for development** menu item to approve the use case. A confirmation dialog opens.

12. Click the Continue button to confirm your choice. The **Status** field changes to **Approved for Development**.

At this point in the lifecycle, the model use case has been created, reviewed for risks, and approved by the various stakeholders. Personas involved are mostly non-technical, from the business user who requested the model to the risk and compliance officer who evaluated it. Next, the model would be developed by teams of data scientists and AI engineers.

The model development steps have their own associated workflows and requirements. For the next part of the lab, you will see how controls can be identified and associated with model risks. You will then examine a model that is under development, to get a closer look at the metrics gathering capabilities of watsonx.governance.

## Controls identification and implementation

Controls are policies and procedures that make sure that risk mitigation responses are performed.

After you identify the risks that occur in your practices, establish controls, such as approvals, authorizations, and verifications. These controls remove, limit, or transfer these risks.

Controls provide either prevention or detection of risks. Controls are associated with tests that ensure that a control is effective. For example, AI models trained with unfairly biased training data frequently exhibit the same unfair bias during their operation.

### 1. View the controls the inventory

Controls are treated as objects in the governance console, and contained in the inventory with questionnaires and assessments.

1. Click on the **menu button** in the upper left to open the menu.
2. Click on the **Assessments** menu item to expand it.
3. Click on the **Controls** menu item. The **Controls** tab opens.

Two sample controls have been created in the environment:

- Screen training data for unfair bias
- Audit data security

Next, you will go through the process of creating a new control similar to the two sample controls. You do not need to save your control or provide any specific details in the form.

4. Click on the [New](#) button to create a new control. The [New Control](#) tab opens.

5. Give your control a name such as *Sample Control*.

6. Set the [Status](#) of the control to [Approved](#).

7. Set the [Control Owner](#) to your user ID.

The screenshot shows the 'Create new control' interface. On the left, the 'General' tab is selected. It contains fields for 'Name' (with value '\_CON\_00000005'), 'Description' (with value 'Sample Control'), 'Status' (set to 'Approved'), and 'Control Owner' (set to 'emartens@us.ibm.com'). Callouts numbered 5, 6, and 7 point to the 'Description', 'Status', and 'Control Owner' fields respectively. On the right, there is a sidebar with the heading 'Create new control' and a note 'Create your new control by filling out the necessary fields.' Below this, it says '1 item requires attention.' and lists 'All Key Items (4)' with 'Name', 'Description', and 'Control Owner' checked, and 'Risk' unchecked.

8. Take a moment to explore and set the values available for [Classification](#), [Control Method](#), [Control Type](#), [Domain](#), and [Frequency](#).

9. Click on the [Select Risk](#) button. The risk selection dialog window opens.

The screenshot shows the 'Select Risk' dialog window. It features several dropdown menus: 'Classification' (Not Determined), 'Control Method' (Automated), 'Control Type' (Detective), 'Domain' (selected items), and 'Frequency' (Monthly). The 'Domain' dropdown is highlighted with a red border. At the bottom right of the dialog is a blue button labeled 'Select Risk' with a numbered callout 9. To the right of the dialog, there is a sidebar with the heading 'Create new control' and a note '1 item requires attention.' and a list of 'All Key Items (4)' with 'Name', 'Description', and 'Control Owner' checked, and 'Risk' unchecked.

Note that the list of risks available in the dialog has been populated from the risks identified earlier in the use cases, or in the master risk library. You can use this screen to assign your control as a process to mitigate a potential risk.

10. When you are finished exploring, either click on a risk to select it and click the [Done](#) button, or click the [Cancel](#) button.

11. Click on the [Cancel](#) button to close the tab and return to the list of controls. Alternately, you can click on the [Save](#) button to save your new control.

## 2. See controls associated with risks

Next, you will explore a control that has been associated with a risk in a use case.

1. Click on the **menu button** in the upper left.
2. Click on the **Inventory** menu item to expand it.
3. Click on the **Use Cases** menu item. The **Use Cases** tab opens.

The screenshot shows the IBM Watsonx Governance console interface. At the top, there's a header with a user icon and some status indicators. Below the header is a navigation bar with several items: 'IBM watsonx', 'Organization', 'Assessments', 'Inventory' (which is highlighted with a blue box and has a circled '2' above it), 'Use Cases' (which is also highlighted with a blue box and has a circled '3' below it), 'Models', 'Model Versions', and 'Model Deployments'. To the right of the navigation bar are various filters like 'Filter By: Criticality', 'Workflow Name', 'Stage', 'Type', 'Stack By: Type', and 'View By: Week'. The main content area shows a list of use cases, with one entry 'Fraud detection emartens' visible.

4. Locate the **Credit risk emartens** use case from the list and click on it. The use case details screen opens.
5. Scroll down to the **Risk** section of the use case and locate the **Risk** table containing all of the identified risks.
6. Locate the **Data bias** risk and click on it. The risk details tab opens.

This screenshot shows the detailed view of the 'Fraud detection emartens' use case. On the left, there's a sidebar with 'Task', 'Activity', 'Admin', and 'Security Performance Monitoring' options. The main area is titled 'Risks' and contains a table with three rows of data. The first row, circled with a '6', is for 'Data bias (Copy of MOD\_0000000\_RIS\_00000001)' with a note: 'Historical, representational, and societal biases present in the data used to train and fine tune the model'. The second row is for 'Data poisoning (Copy of MOD\_0000000\_RIS\_00000002)' with a note: 'Data poisoning is a type of adversarial attack where an adversary or malicious insider injects intentionally corrupted, more'. The third row is for 'Over or under reliance (Copy of MOD\_0000000\_RIS\_0000030)' with a note: 'When a person places too little or too much trust in an AI model's guidance.' To the right of the table, there's a 'Copy from Library' button. Further right, there's a panel for 'Initial Approval' with fields for 'Stage' (set to 'Initial Approval (Awaiting use case approval)'), 'Due Date' (set to '2/18/2025'), and 'Tags' (empty). Below the approval panel, there's a section for 'Initial Approval' with instructions: 'Please review the initial details related to the use case as captured by the Use Case Owner. Use the Actions button to Return to owner,' followed by a 'more' link.

7. Scroll down to the **Related Content** section of the screen, and note that the **Mitigating Controls** tab contains a table with one control.
8. Click on the control in the table. The control information tab opens.

At this point, you could kick off an assessment of the control to verify that it was being performed to mitigate the risk. You have now seen how mitigating controls can be defined by an organization, and associated with risks that they are meant to mitigate. You then saw how those controls can be attached to use cases to ensure that stakeholders are taking the required steps to mitigate the risks.

The governance console gives clients the flexibility to make these controls part of approval workflows, and to build reports and metrics based on the completion of the controls.

In the next section, you will explore how watsonx.governance can evaluate and monitor AI and machine learning models.

## Monitoring and issue management

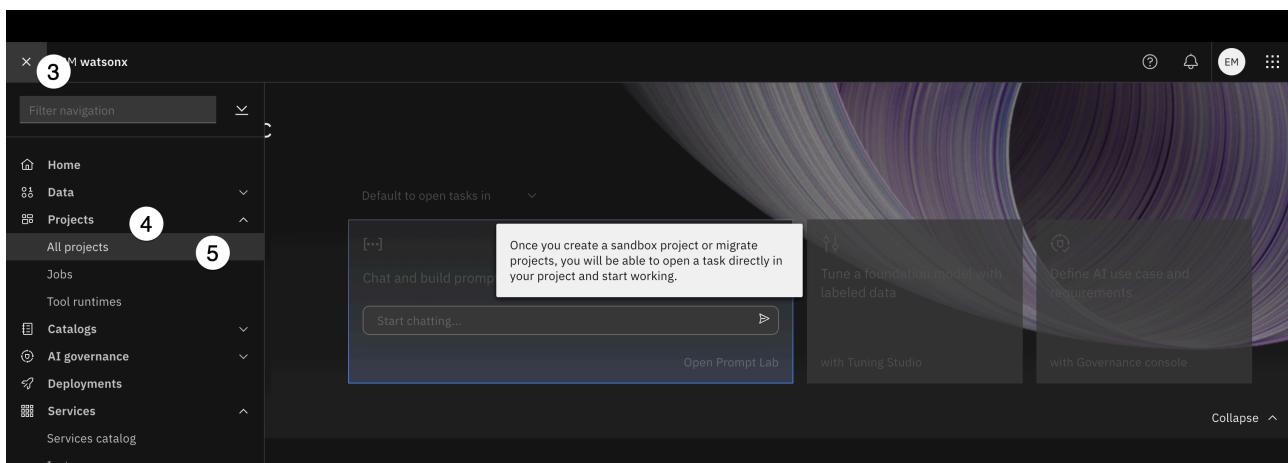
In this section, you will explore the various metrics automatically gathered by the watsonx.governance solution. You will work with a project in the watsonx development environment, from the point of view of a model developer or AI engineer. However, the same functionality is available for models developed and deployed in third-party platforms such as Microsoft Azure or AWS SageMaker.

## 1. Explore the development environment

1. Click on the [menu icon](#) in the upper left.
2. Click on the [IBM watsonx](#) menu item. You will return to the watsonx home screen.



3. From the watsonx home screen, click on the [menu icon](#) in the upper left to open the menu.
4. Click on the [Projects](#) menu item to expand it.
5. Click on the [All projects](#) menu item. The [Projects](#) screen opens.



6. Locate the [Credit risk development](#) project from the list and click on it. The project details screen opens.

IBM watsonx uses projects to allow data scientists, data engineers, and subject-matter experts to collaborate on data science projects. They can contain a huge variety of assets, including (but not limited to):

- Machine learning models
- AI prompt templates
- Jupyter notebooks
- Database connections
- AutoAI experiments
- SPSS Modeler flows

7. Click on the [Assets](#) tab of the project.

This particular project contains a comma-separated value (CSV) file of data that has been used to train a credit risk model with AutoAI, IBM's auto-machine learning tool for rapid model prototyping. The AutoAI

experiment is also contained in the project, as is the output model.

8. Click on the [Credit risk experiment - P5 XGB Classifier - Model](#). The model factsheet opens.

The screenshot shows the WatsonX Assets interface. The top navigation bar includes 'Projects / Credit risk development', 'Overview', 'Assets' (highlighted with a black box and a circled '7'), 'Jobs', 'Manage', and various icons for filtering and saving. Below the navigation is a search bar with 'Find assets'. The main area displays a table titled 'All assets' with columns for 'Name', 'Last modified', and a 'More' icon. The table contains three entries: 'credit\_training\_data.csv' (modified 1 hour ago by you), 'Credit risk experiment' (AutoAI experiment, modified 1 hour ago by you), and 'Credit risk experiment - P5 XGB Classifier - Model' (Machine learning model from AutoAI, modified 37 minutes ago by System). A callout circle labeled '8' points to the model asset.

Model factsheets collect all the relevant data around a particular model, including both evaluation metrics and metadata. Factsheets are created when the model developer initiates model tracking. In the [day 2 lab](#) you will receive hands-on experience with this process. For now, understand that when you turn on tracking for the model, you associate it with an AI use case like the ones you created earlier in this lab.

In a real workflow, the stakeholders would create a use case request, and go through the AI use case request process that you completed earlier. Recall that at the end of that process, you approved the model for development. At that point, data scientists and AI engineers would begin developing the model using tools like the watsonx project, prompt lab, or similar tools from Azure, AWS, or other vendors.

9. Scroll down to the [Development](#) section, or use the quick links in the left-hand navigation. Take a moment to review the captured metadata, including the model author and link to the data used to train the model.

The screenshot shows the 'Development' factsheet for the 'Credit risk development' project. The left sidebar lists sections like Governance, Model, Development (selected), and Validation. The main content area has a title 'Development' and a sub-section 'Credit risk development' with ID 'c06ac4b0-375b-4079-82ff-2956bc8a5c38'. It includes a 'Description' field ('Develop a credit risk model'), 'About this asset' details (Created by Eric Martens, Created February 12, 2025 at 09:42:33 PM, Asset name 'Credit risk experiment - P5 XGB Classifier - Model'), and a 'Training evaluation' section. The 'Training evaluation' section shows a table with three rows: 'Training data source' (highlighted with a red box) containing 'credit\_training\_data.csv', 'Asset type' (Data asset), and 'Source type' (CSV). A callout circle labeled '9' points to the 'Training evaluation' section.

10. Scroll down further to the [Training evaluation](#) section of the page, and note that quality metrics data generated by AutoAI during the creation of the model has been saved in the factsheet. These metrics represent the ones calculated using holdout data during model creation. In particular, note the [Roc auc](#) score of 0.84.

Metric	Training data	Holdout data
Accuracy	0.8007791	0.775
Average precision	0.9076312	0.49300915
Balanced accuracy	0.74765736	0.7090683
F1	0.85871994	0.8432056
Log loss	-0.44427747	-0.4481161
Precision	0.81742966	0.7908497
Recall	0.90440565	0.9029851
<b>Roc auc</b>	<b>0.8423995</b>	<b>0.84271824</b>

**Area under ROC** is a standard measurement of quality for classification models, particularly binary classification models such as this credit risk example.

- Finally, note that you can add attachments to the factsheet. These could include additional requirements, reviews, training data information, or any other details that should be associated with the model.

Next section, you will take a closer look at some of the metrics that have been calculated by the monitoring service.

## 2. Examine metrics from the monitors

In this section, you will look at metrics for the model. Depending on regulations, risks, and model types, different metrics such as quality and fairness will have different thresholds they must meet to remain in compliance. These thresholds can be configured at the model deployment level, allowing organizations to fully customize their metrics reporting and compliance safeguards.

For example, a global company may have one copy of their credit risk model deployed in the United States, where it needs to meet a fairness threshold of 80%. They could then deploy a second copy of the same model in a different region that had a different fairness threshold, and configure monitoring for each model to generate alerts if the fairness dropped below the threshold in that region.

The monitors for this model have already been configured, and an evaluation has been run. Again, in the [day 2 lab](#), you will get an opportunity to perform this configuration.

- Use the left-hand navigation or scroll down to the [Evaluation results](#) section of the factsheet, and examine the [Area under roc](#) metric.

Metric	Value
True positive rate (TPR)	0.55
Precision	0.63
Accuracy	0.76
Area under roc	0.70
F1-measure	0.59
Logarithmic loss	0.47

As you can see from the chart, the minimum threshold for this measurement has been set to 0.8; however, in the most recent evaluation, the model scored 0.70, meaning that the threshold has been violated. The metric shows in red with an alert badge to note the violation. This score is lower than the score calculated

when training the model (0.84), showing that the evaluation data set was somewhat different than the holdout data used to train the model.

2. Hover your mouse over the **information icon** in the **Quality** section to see extra information about the evaluation.

The screenshot shows the 'Evaluation results' dashboard. On the left, a sidebar lists 'Development' and 'Validation' sections, with 'Evaluation results' selected. The main area is titled 'Quality' and displays four metrics with red information icons: True positive rate (TPR) at 0.55, Area under roc at 0.70, Precision at 0.63, and F1-measure at 0.59. A callout box points to the 'View details' link next to the information icon for the TPR metric. The top right shows the date 'February 12, 2025 at 10:40:03 PM' and 'Records: 100'. The bottom right contains asset details like Name ('Credit risk experiment - P5 XGB Classifier - Model'), Type ('wml-hybrid\_0.1'), and Software specification ('hybrid\_0.1').

Note that the evaluation only used 100 records, which is a fairly small sample size. A data scientist might recommend a new evaluation with more data to see how that affects the metrics before trying to retrain the model.

3. Click on the **View details** link to the right of the **information icon**. A new browser tab opens into the monitoring service (formerly Watson OpenScale) dashboard. This dashboard offers more in-depth information on the various quality metrics, including a confusion matrix.
4. Click on the **Fairness** tab in the dashboard. The **Fairness** evaluation screen opens.

The screenshot shows the 'Fairness' evaluation screen. The top navigation bar includes 'Dashboard / Credit risk candidate /' and tabs for 'Fairness', 'Quality' (which is active), and 'Drift v2'. Below the tabs is a 'Quality' section with a note about comparing automatically produced classification against a reference classification. A table lists various fairness metrics with their values and threshold violations:

Area under ROC	Area under PR	Accuracy	True positive rate (TPR)	False positive rate (FPR)	Recall	Precision	F1-Measure	Logarithmic loss	Brier score	Matthews correlation coefficient
0.7 ① 0.1 lower threshold violation	0.56 ① 0.24 lower threshold violation	0.76 ① 0.04 lower threshold violation	0.55 ① 0.25 lower threshold violation	0.14	0.55 ① 0.25 lower threshold violation	0.63 ① 0.17 lower threshold violation	0.59 ① 0.21 lower threshold violation	0.47	0.53	0.42 ① 0.38 lower threshold violation

A 'Confusion matrix' button is visible at the bottom left.

The administrator has chose to monitor two different variables for fairness in this example: sex and age. The **disparate impact** calculation has been used to calculate fairness. Again, thresholds for acceptable fairness can be set for each deployment of the model, allowing organizations to maintain compliance wherever they do business.

5. Take a moment to review the details on the **Fairness** screen. You can click on the **View calculation** link to see how disparate impact is calculated, or toggle between a percentage of positive outcomes and the raw number of positive and negative outcomes.
6. Click on the **Monitored attribute** dropdown and select **Age**. The graph changes to show the fairness metric for the **age** feature, which is also within acceptable limits.

The screenshot shows the Fairness tab of the WatsonX Governance interface. The 'Monitored attribute' dropdown is set to 'Sex'. The 'Fairness metric' dropdown is set to 'Disparate impact'. The 'Data set' dropdown is set to 'Balanced'. Below these settings, the 'Disparate impact' score is displayed as 88%, with a note indicating it's for the female group. A 'View payload transactions' button is present. At the bottom, there are options to 'View percentage' or 'View count'.

Note that, in most cases, age and sex would not be features of a model, as that would increase the chance of unfair bias. The watsonx.governance solution allows for indirect bias monitoring, where the data is submitted with the model records as metadata, and not actually sent to the model for evaluation. In this way, organizations can monitor for unfair bias without possibly introducing it into their models.

7. Click on the **Drift v2** tab. The drift metrics appear.

The screenshot shows the Drift v2 tab of the WatsonX Governance interface. It displays historical drift scores for February 12, 2025. The scores are listed as follows:

- Output drift: 0.218
- Prediction drift: 0.0139
- Model quality drift: 0.065
- Feature drift: 0.2249

A time series chart on the right shows a vertical timeline with colored dots representing the drift values at different points in time. The chart has a legend indicating three categories: blue, orange, and red. The last evaluation date shown is Feb 12, 2025, 10:41 PM.

Drift is a measurement of how changes in data over time affect model accuracy. For example, models that predict demand for toilet paper and sanitizing wipes may have been accurate in January 2020, but radically increased demand brought on by the Covid-19 pandemic was not anticipated, and immediately impacted the model's quality. Not all cases are this extreme; construction of a new factory in a small town would slowly bring more people to the area, driving up demand for housing, schooling, and other goods and services, and affecting any models trying to make predictions about that demand.

An in-depth discussion of drift is beyond the scope of this lab. Know that watsonx.governance can monitor drift continuously for models in production, allowing stakeholders to see and react to quality issues like the ones described above in real time without the need for gathering additional feedback data to take quality metrics.

8. Return to the browser tab showing the model factsheet, and scroll down in the **Evaluation results** section to the **Drift\_v2** and **Fairness** sections.

**Drift\_v2**

**Payload**

- Model quality drift: 0.07
- Output drift: 0.22
- Feature drift: 0.22
- Prediction drift: 0.01

**Fairness**

**Sex**

- Disparate impact: 88.31%

**About this asset**

**Name:** Credit risk experiment - P5 XGB Classifier - Model

**Description:** No description provided.

**Asset Details:**

- Type: wml-hybrid\_0.1
- Model ID: 32dda205-b2c1-44...
- Software specification: hybrid\_0.1
- Hybrid pipeline software specifications: autoai-kb\_r124.1-py3.11

**Tags:** Add tags to make assets easier to find.

Note that the same metrics you viewed in the monitoring service dashboard are available here, with slightly less detail, as this view is meant to be more of an overview of the metrics.

In this section, you explored the different metrics available to AI engineers and model developers. These metrics can be calculated on models in development, as well as continuously monitored for models in production deployment. Furthermore, customizable metrics thresholds allow organizations to ensure that all of their models are in compliance with regulatory and industry standards.

The model you examined is meeting fairness standards, indicating that it is not unfairly biased. However, it is falling short in several quality metrics, indicating that the data scientists and AI engineers likely have more work to do before they can advance the model as a candidate for production deployment.

## Dashboards and reporting

In this section, you will see how the watsonx.governance console provides an enterprise-level view of all the use cases and models in development, and see how the metrics generated in the previous section appear in the governance console dashboard. This section is once again performed from the point of view of a risk officer or business stakeholder, as opposed to the data scientists and AI engineers from the previous section.

### 1. View metrics for the credit risk use case

1. Scroll back up to the [Governance](#) section of the factsheet, and click on the link beneath [Asset record](#). The governance console opens to the entry for the model.

**Credit risk emartens**

Under Development | System | Deloitte Lab Inventory | Medium |  
e80e5c23-94b2-423f-a93a-19511807927f

Description  
Predict risky credit

Approach Version  
Default approach 1.0.0

A default approach for tracking your AI assets.  
00000000-0000-0000-000000000000

Asset record  
**Credit risk experiment - P5 XGB Classifier - Model**

Asset record status  
Proposed

Last modified  
10 hours ago by System

Created on  
Feb 12, 2025 by Eric Martens

Note that the **Candidate Status** of the model is set to **Model Candidate**, and the **Model Status** is set to **Proposed**. These values reflect that the model is in development and has not been approved for deployment to production systems yet.

2. Click on the [home](#) icon to return to your home screen.

3. Click on the [Home](#) tab to view your home dashboard.

Welcome, Eric Martens!

Last successful login 2/12/2025, 7:40 PM

Home My Tasks (3) Subscription Tasks (0) Oversight Tasks (0)

Models by Department  
2

Change Requests by Status  
0

My Active Change Requests  
0

Change Requests in Process

Model Inventory  
Use Cases by Risk Level  
3

Models by Risk Tier  
2

Use Cases by Lifecycle Phase  
3

- Awaiting Use Case Appr
- Approved for Developm
- Under Development**
- Proposed
- Rejected
- Awaiting Development
- Developed
- Ready for Validation
- Validation Complete

The home dashboard contains several different panels showing a breakdown of all the models currently being governed, along with useful links to different regulations and the [IBM AI Risk Atlas](#). This dashboard can be fully configured with different views for different users or profiles. Recall that you set your profile to [watsonx-governance MRG Master](#) earlier, but that you have access to several other profiles.

4. Click on the [avatar](#) icon in the upper right to open the user menu.

5. Click on the [Change profile](#) menu item to open the profile selection dialog.

6. Take a moment to switch back and forth between some of the available profiles, and note how the dashboard changes based on your selected role. Also recall that different roles will see different views of the same object; for example, the *OpenPages Modules Master* will not have access to fill out the mitigation strategies for various risks, as that capability is not part of their profile.

7. When you are finished, return to the *watsonx-governance MRG Master* profile and click on the **Under Development** portion of the **Use Cases by Lifecycle Phase** graph. The **Use Cases** tab opens, with a filter to show only the use cases that are currently under development.

8. Locate the [Credit risk emartens](#) use case from the table and click on it. The use case details screen opens.

9. Scroll down to the **Performance Monitoring** section. Here, the metrics created by the model evaluation you saw in the previous section are available for risk manager and business stakeholders in a variety of different views. Take a moment to examine the different charts, and note that you can click on many of the indicators for more information on a particular item.

10. Scroll down further to the **Metrics in Breach** table and click on the [All Metrics](#) tab.

Name	Description	Evaluation Category	Evaluation Sub-Category	Value	Breach Status	Tags
Accuracy GlobalCorp > Corporate Finance	Watson Studio Notebook metric for Accuracy		Accuracy	0.8007791	Not Determined	<span style="color: green;">Green</span>
Average precision GlobalCorp > Corporate Finance	Watson Studio Notebook metric for Average precision		Average precision	0.9076312	Not Determined	<span style="color: yellow;">Yellow</span>
Balanced accuracy GlobalCorp > Corporate Finance	Watson Studio Notebook metric for Balanced accuracy		Balanced accuracy	0.74765736	Not Determined	<span style="color: yellow;">Yellow</span>
F1 GlobalCorp > Corporate Finance	Watson Studio Notebook metric for F1		F1	0.85871994	Not Determined	<span style="color: yellow;">Yellow</span>
Feature_drift_subscription GlobalCorp > Corporate Finance	Feature drift metric of Drift_v2	Other	Feature drift	0.2249	Red	<span style="color: red;">Red</span>

Stage: Under Development  
Due Date: 2/18/2025  
Tags: No tags have been added yet.  
Use Case under development: The use case is currently being developed. When the development is completed, use the Actions menu to mark development complete. Alternatively, the development project can be closed.

Here, each metric is visible along with its value and breach status (red, yellow, green, or not determined).

The integration between the monitoring service and the governance console allows for model metrics to be calculated and instantly updated and provided to all stakeholders. This can provide dramatic time savings, as it eliminates the need for data scientists to build reports on model performance for stakeholder review. It also eliminates the possibility of human error, and ensures that reviewers have the most up-to-date information at all times.

## 2. View metrics for generative AI

The watsonx.governance solution offers a comprehensive set of metrics for generative AI as well.

1. Click on the [home](#) icon to return to your home tab.
2. Scroll down to the [Models by Provider](#) section on the left, and click on the number beneath [Foundation Models](#). The [Models](#) tab opens, with a filter set to show only foundation models.

IBM WatsonX | Governance console

Welcome, Eric Martens!  
Last successful login 2/14/2025, 3:05 PM

Home My Tasks (3) Subscription Tasks (0) Oversight Tasks (0)

Models by Provider

Foundation Models: 1  
Developed Models and Prompts: 3

Oversight Tasks (0)

Compliant Non-compliant

New Use Case +

SR 11-7 Information  
E-23 Information

Deployments

Deployments by Status: 2  
Production Deployments: 0

Issues by Status: 0  
No data available

3. The table contains an entry for OpenAI's [GPT-3.5-turbo](#). Click on it. The model details open in a new tab in the governance console.
4. Scroll down to the [Associations](#) section, and note that there is an associated use case. This AI use case contains a model that is using the GPT-3.5-turbo large language model via a detached prompt template. Detached prompt templates will be covered in greater detail in the [day 2 lab](#).

The [Associates](#) section provides an easy method to determine all the different use cases that are utilizing a particular model, information that would be highly useful for organizations looking to determine the extent of their usage and how any outages, updates, or other issues may impact them.

- Click on the [Resume summarization use case](#) from the [Related Use Cases](#) table. The tab for the use case opens.

The screenshot shows the 'Associations' table in the 'Related Use Cases' section. A row for 'Resume summarization emartens' (GlobalCorp > Human Resources) is selected, highlighted with a black circle containing the number 5. The table includes columns for Name, Purpose, Description, Status, Risk Level, and Third Party Link. The status is 'Approved for Development' with a 'High' risk level, and the link points to a specific URL. The right panel displays the 'Data Check' workflow stage, which is currently set to 'Data Check (Data)' with a due date of 'Not Set'. It also shows a note about ensuring foundation model quality and includes a 'more' link.

- Scroll down to the [Performance Monitoring](#) section of the use case. As with the previous model, you can click on any of the metrics to get more in-depth information on the score and the status. The [Metrics in Breach](#) table also provides a view of all metrics that fell below the acceptable thresholds.

The use case view shows a summary of all models associated with a particular use case, but you can also get more information on a particular deployment.

- Scroll down further to the [Related Models](#) section of the use case. In the [All Models](#) tab of the table, note that there are entries for both the prompt and the source large language model.
- Click on the [Deployments](#) tab of the table to get a listing of all deployed models associated with the use case.
- Click on the [Resume summarization deployment](#) entry in the table. A new tab for the deployment opens in the governance console.

The screenshot shows the 'Deployments' table in the 'Related Models' section. A row for 'Resume summarization deployment' (GlobalCorp > Human Resources) is selected, highlighted with a black circle containing the number 9. The table includes columns for Name, Description, Environment, Deployment Status, and Third Party Link. The status is 'Deployed'. The right panel displays the 'Use Case general view' with a detailed description of what a use case is and its purpose, along with sections for 'All Key Items' and filter options for 'Purpose', 'Risk Level', and 'Use Case Type'.

The entry for the deployment contains metrics from the most recent evaluations, links back to the parent use case, and other metadata.

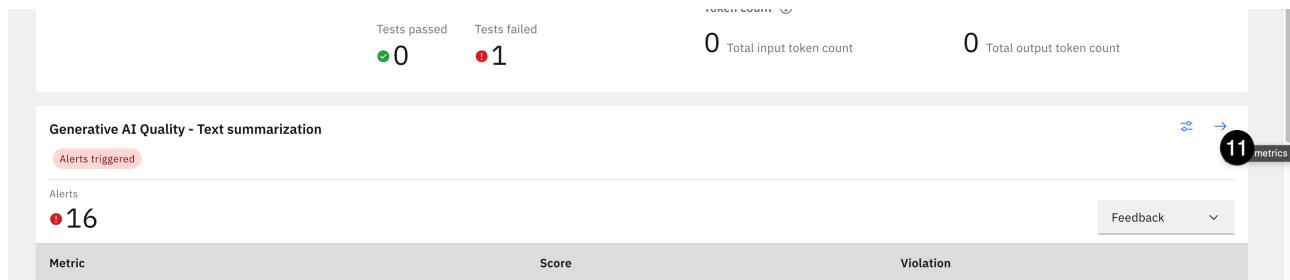
- In the [General](#) section of the page, click on the [Third Party Link](#). A new browser tab opens to the model's factsheet in the watsonx interface.

The screenshot shows the 'General' section of the deployment details. It includes fields for Name (Resume summarization deployment), Description, Deployment Status (Deployed), Initial Implementation Date, Business Owner, Delegate, and a 'Monitored with watsonx.governance' field set to 'Yes'. The 'Third Party Link' field contains a URL. The right panel displays the 'Stage' and 'Tags' sections, and a 'Deploy Model' section with instructions for performing activities like retrieving model assets and deploying them.

Recall that factsheets metrics and metadata for the model, and are automatically synchronized with the monitoring service and the governance console. The current view is designed for data scientists and AI

engineers, as opposed to the governance console, which is targeted more towards business users and risk managers.

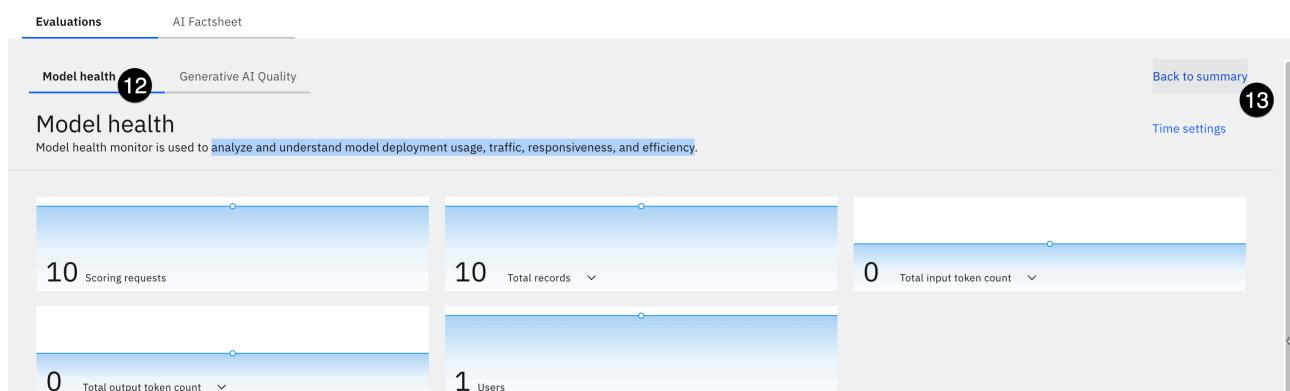
11. In the **Generative AI Quality** section, click on the arrow icon for a better view of the available metrics.



12. Take a moment to explore the different metrics for quality, content analysis, HAP, and more by clicking on the different sections to expand them. When you are finished, scroll back to the top of the screen and click on the **Model health** tab.

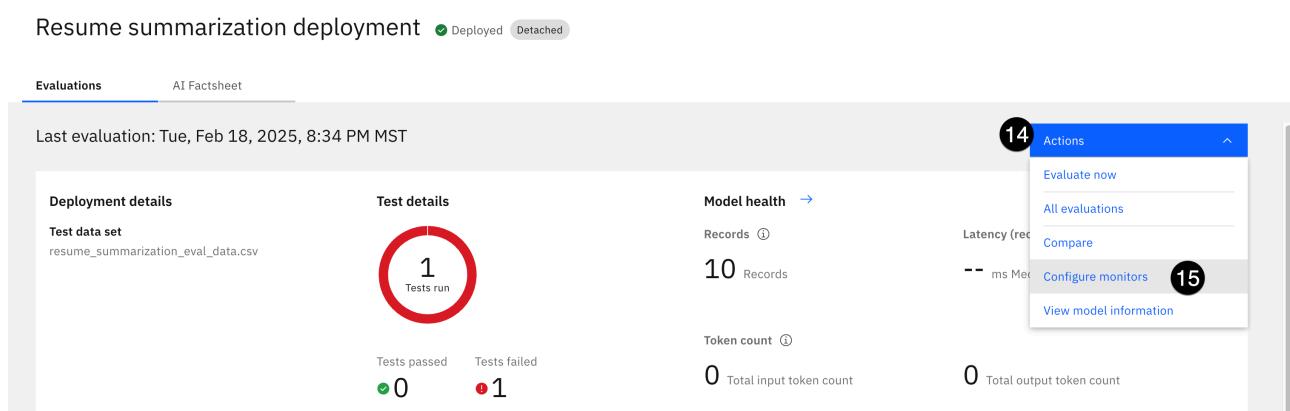
The model health metrics provide information on model deployment usage, traffic, responsiveness, and efficiency.

13. When you are ready, click the **Back to summary** link to return to the previous tab.



14. Click on the **Actions** button to open the **Actions** menu.

15. Note the different actions available, including running evaluations, comparing two models, and seeing a full evaluation history. Click on **Configure monitors**.



16. Click on the **Generative AI Quality** item from the menu on the left.

Take a moment to examine the tiles on the right, which include the different metrics categories available and show the minimum thresholds for each metric. Administrators with proper authority can use this screen to change the thresholds based on legal or industry requirements, which will then affect whether the metrics show on the factsheet and in the governance console as in breach.

Note that any changes you make here will affect all other lab users, so please do not save any changes to the metrics.

- When you are finished, close the current browser tab and return to the tab with the governance console.

### 3. View other use cases

At this point in the lab, take a moment to review some of the use cases created by your colleagues, and the various risks that they have identified.

- Click on the [home](#) icon to return to your home screen.
- Locate the [Model Inventory](#) tile, and click on the number to the left of the [Use Cases by Risk Level](#) graph.

The use cases that all of your colleagues have built as they have gone through the lab will be visible, along with their associated risk levels.

- Click on some of the use cases your colleagues have created, and explore the different risks that were identified. If you wish you can even click on the risk identification questionnaire and the applicability assessment for the use case to see the answers they provided.

## Conclusion

In this lab, you got experience with watsonx.governance from the perspective of daily users -- risk managers, business stakeholders, data scientists, and AI engineers. You saw how to identify a use case in which an organization can use AI to solve a business problem. You learned how that request can progress through a defined workflow, using questionnaires to identify potential risks and compliance issues.

Next, you explored how watsonx.governance fits into the model development process, capturing model facts and metadata in a central location. You saw some of the different metrics available in watsonx.governance for predictive models, and how they can be evaluated against customizable thresholds to ensure models meet regulatory and industry standards.

Finally, you saw how those metrics are automatically updated in the governance console, providing up-to-date information for risk managers and other stakeholders.