



watsonx.governanceTM



watsonx.governance Hands-On Workshop

*Introduction to AI Governance via **Building**, **Deploying**, and **Managing** AI models.*

For more information, visit us at: <https://www.ibm.com/products/watsonx-governance>

Notices and Disclaimers

This guide contains proprietary information which is protected by copyright. No part of this document may be photocopied, reproduced, or translated into another language without a legal license agreement from IBM Corporation.

© Copyright IBM Corporation 2025.

The information contained in these materials is provided for informational purposes only and is provided AS IS without warranty of any kind, express or implied. IBM shall not be responsible for any damages arising out of the use of, or otherwise related to, these materials. Nothing contained in these materials is intended to, nor shall have the effect of, creating any warranties or representations from IBM or its suppliers or licensors, or altering the terms and conditions of the applicable license agreement governing the use of IBM software. References in these materials to IBM products, programs, or services do not imply that they will be available in all countries in which IBM operates. This information is based on current IBM product plans and strategy, which are subject to change by IBM without notice. Product release dates and/or capabilities referenced in these materials may change at any time at IBM's sole discretion based on market opportunities or other factors and are not intended to be a commitment to future product or feature availability in any way.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

IBM, the IBM logo and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at ibm.com/legal/copytrade.shtml

Other company, product and service names may be trademarks or service marks of others.

Table of Contents

Notices and Disclaimers	2
Introduction to watsonx.governance	4
Introduction to your lab scenario	5
Phase 1: Plan	5
Create an AI Use Case	5
Risk Assessment	11
EU AI Act applicability Assessment	16
Submit for stakeholder review	18
Create a Prompt Template	23
Model Parameters	28
Saving Prompt	29
Evaluating your prompt template	31
Evaluation Results	34
AI Factsheet	36
Govern your prompt template	39
Tracking Prompt Template in a Use Case	42
Update your prompt template	47
Evaluation Studio	49
Phase 4: Deployment	54
Deployment Space	54
Managing Assets in production	57
Drift Monitoring (Optional – feel free to progress to the Governance Console section)	62
Governance Console	69
Optional steps: Navigating the UI of the Governance Console	73
Promoting your prompt to Validation	92
Deploying the prompt	95
Activate prompt monitoring (Evaluation).....	97

Introduction to watsonx.governance

watsonx.governance is a Modern AI Governance toolkit that provides a sleek and intuitive User Interface. Whether Users are looking for responsible AI workflows, or to reduce costs and comply with regulations, watsonx.governance provides the capabilities to empower everyone in the organization with the insights needed to positively impact governance.

This workshop is designed to give you an opportunity to discover the ease of use of watsonx.governance, IBM's AI Governance solution that enhances the efficiency and capabilities of AI developers, users, policymakers and ethicists alike.

In this workshop, you will be creating a Use Case which uses AI to summarise a Car Insurance Claim. The scenario will allow you to experience the following capabilities in watsonx.governance:

- Creation of a Use Case including automated workflows
- Performing Risk and Compliance Assessments
- Gaining approvals for the sign off to Development
- Development of a prompt template
- Test your prompt, try different parameters and models.
- Evaluate your prompt against gold standard summaries.
- Track your prompt through its lifecycle within the AI use case. This includes version control.
- Promote and evaluate your prompt template in a pre-production deployment space.
- Managing the prompt in production.

Introduction to your lab scenario

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.

Phase 1: Plan

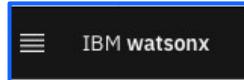
We define a Use Case as a collection of information to solve a business problem, in this case an Insurance Claim Summarisation. In practice, an organization would identify a business need for an AI model, and create a use case to track that effort.

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 planning, design, development, and production phases.

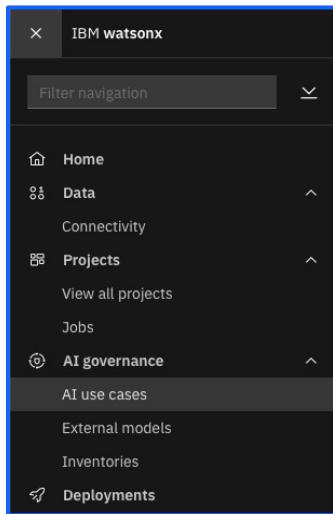
In the planning phase, you will use pre-configured workflows to populate the necessary information about the Use Case itself.

Create an AI Use Case

1. Please select the hamburger menu on the top left.



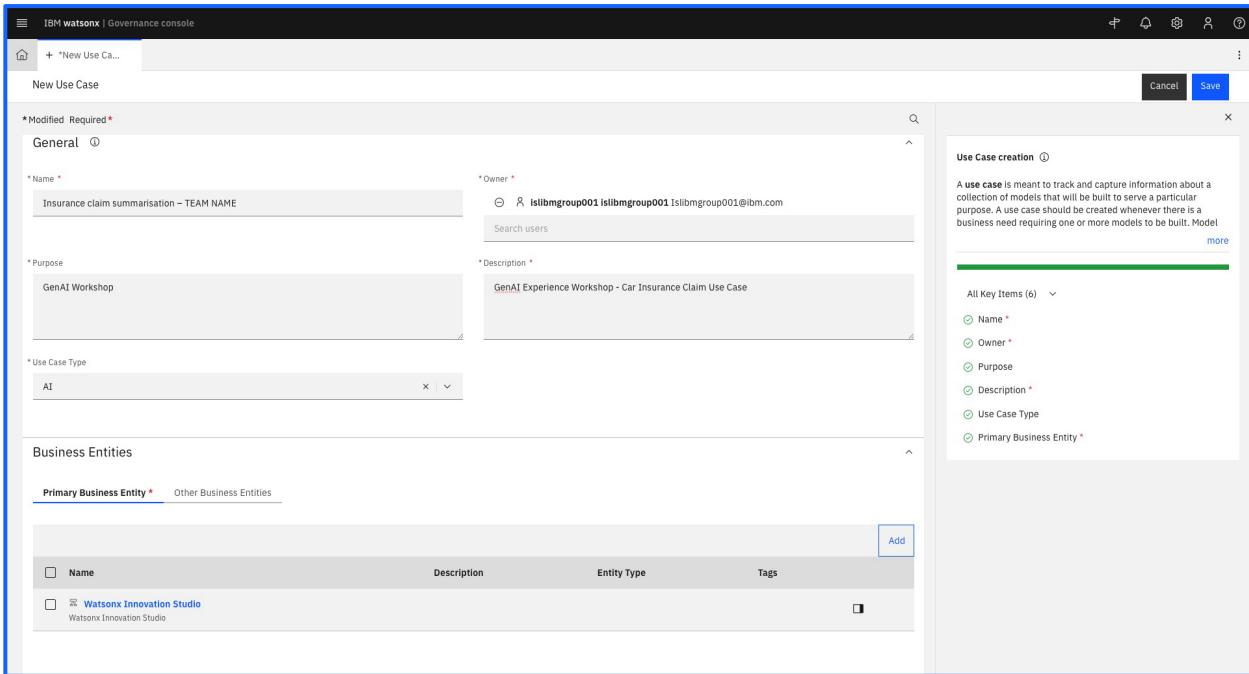
2. Select **AI use cases** from the menu.



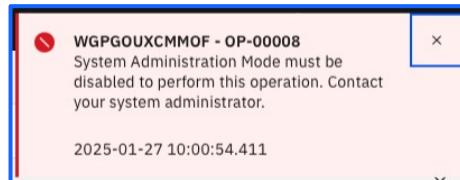
3. Create a new AI use case. At this point you will be redirected to the watsonx.governance console however once the steps are completed the Use Case will also be visible in your list of

AI Use Cases screen. This is an example of the bi-directional flow between the Evaluation/Monitoring and AI Factsheet modules and the Governance Console

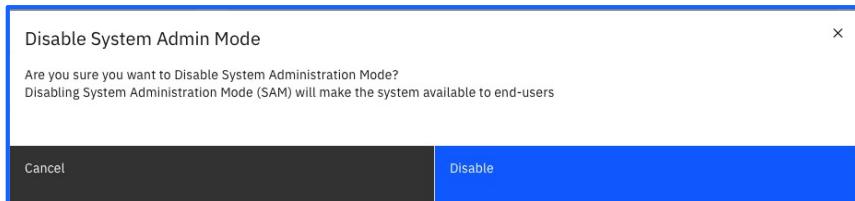
- 
4. Fill out the Use Case using the following parameters:
 - Give it a **Name** such as “Insurance claim summarisation – *(insert your team name)*”
 - Add the **Owner** from the sample accounts, i.e. islibmgroup001
 - **Primary Business Entity** set “Watsonx Innovation Studio”
 - Click **Save**
 5. Next, fill out details which are applicable to the Use Case itself, e.g.:
 - **Description** set “GenAI Experience Workshop - Car Insurance Claim Use Case” • **Use Case Type** set “AI”
 - *Any others?*



NOTE:If you receive the following error:



Then from Setting disable “**System Admin Mode**”



Note: you may also get an error indicating someone else has updated the Use Case/Object. If you do, please refresh the screen.

6. Whilst in Edit mode set the **Technical Owner** to something like “islibmgroup001” as well as the

Stakeholder Departments to “Legal” and hit at the top right of the screen.

IBM Watson | Governance console

*Insurance c...

Use Case
Insurance claim summarisation ⋮ ⌂

Status Proposed Risk Level Low

Cancel Save

Task Activity Admin Security Performance Monitoring

*Modified Required *

General ⌂

Name * Insurance claim summarisation Use Case Type AI Status Proposed

Description GenAI Experience Workshop - Car Insurance Claim Use Case Owner isibmgroup001 isibmgroup001 isibmgroup001@ibm.com Purpose GenAI Workshop

Tags No tags have been added yet.

Use Case Data Gathering ⌂ Please capture all relevant information to this AI use case proposal and then submit using the Action button.

All Key Items (5) ▾

Selected Item isibmgroup001 isibmgroup001 isibmgroup001@ibm.com

Stakeholder Departments

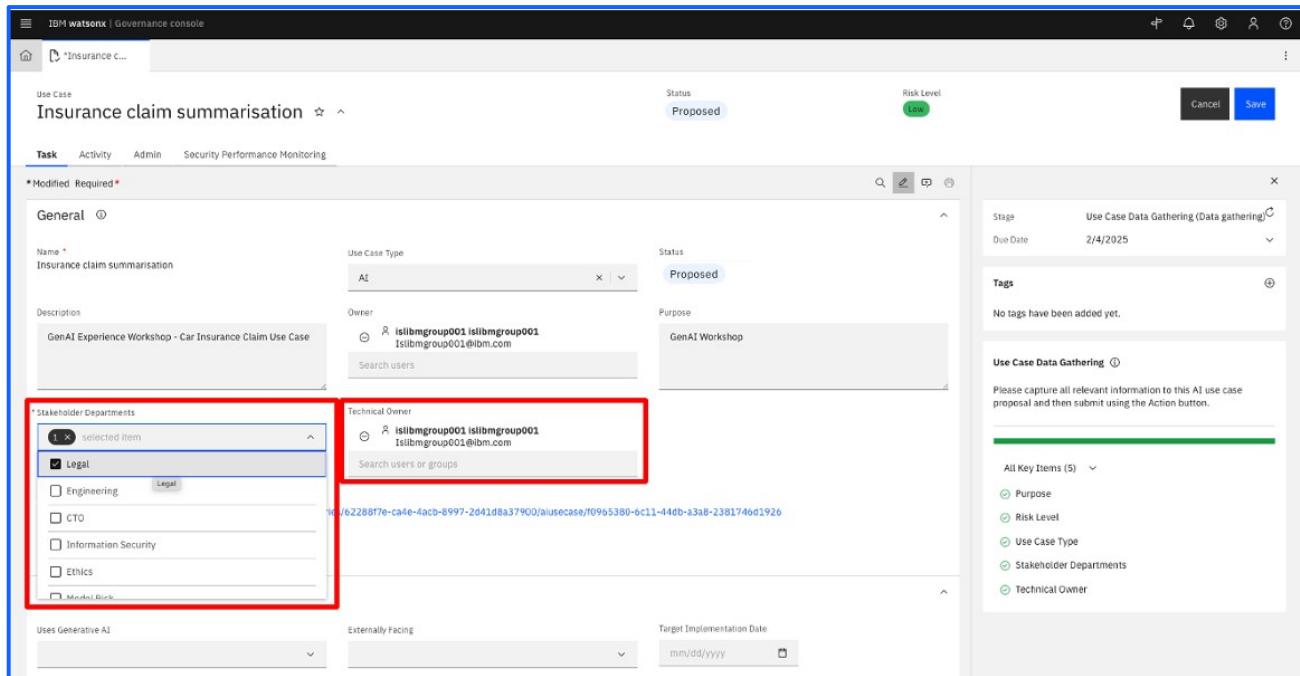
- Legal
- Engineering
- CTO
- Information Security
- Ethics

Technical Owner

isibmgroup001 isibmgroup001 isibmgroup001@ibm.com

Search users or groups

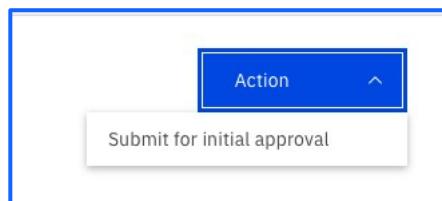
Uses Generative AI Externally Facing Target Implementation Date mm/dd/yyyy



7. The Use case should look like this:

The screenshot shows the IBM Watsonx Governance console interface. At the top, there's a navigation bar with 'IBM watsonx | Governance console', a search bar, and several icons. Below the navigation is a breadcrumb trail: 'Home > Insurance cl... > Use Cases > + New Use Case'. The main area shows a use case card for 'Insurance claim summarisation'. The card includes fields for Name (Insurance claim summarisation), Use Case Type (AI), Status (Proposed), Risk Level (Low), and a detailed description of the use case purpose and owner. To the right of the card is a large modal window titled 'Action'. This modal contains a sub-section titled 'Use Case Data Gathering (Data gathering)' with a due date of 2/4/2025. It also includes sections for 'Tags' (none listed) and 'Use Case Data Gathering' with instructions to capture all relevant information. At the bottom of the modal is a button labeled 'Submit for initial approval'.

8. Now that you have successfully created the Use Case you can **Submit it for Initial Approval** by navigating to the **Actions** button at the Top right of the screen:



The Action button only works when the required information is captured. This is all driven by customisable Workflows which are set up to match an organisations business process'. This is outside the scope of the labs however if you wish to see this, please ask the table facilitator to demonstrate.

9. On the pop-up window select **Continue**:



10. Once the Use case is Submitted for Approval, the Use Case gets updated and under the **Risk** section further down the screen, an **AI Risk Identification Assessment** has been automatically created:

Name	Description	Progress (%)	Tags
<input type="checkbox"/> AI Risk Identification (Insurance claim summarisation)	Watsonx Innovation Studio AI use case risk identification assessment (2025-01-30)		

Risk Assessment

As part of the planning phase, we need to perform a risk assessment of our AI Use Case. The following steps show how an organisation can use a pre-build risk assessment to identify and manage the AI Risks associated with their Use Case.

1. Select the Assessment:

The screenshot shows the IBM Watsonx Governance console interface. The top navigation bar includes links for 'IBM watsonx | Governance console', 'Home', 'Insurance cl...', 'Use Cases', '+ New Use Case', 'AI Risk Ident...', and user profile icons. Below the navigation is a secondary menu with 'Task', 'Activity', 'Admin', and 'Questionnaire' tabs, with 'Questionnaire' being the active tab. On the left, a sidebar shows 'Questions completed' at 0/19, a 'Copy answers' button, and a 'Sections' dropdown set to 'Main Section'. The main content area is titled 'Main Section 0/19' and 'AI Use Case Risk Identification 0/19'. It contains an 'Introduction' section with a note about the goal of identifying AI risks. Below it are two questions: 'Problem statement' (labeled 1) asking for a description of the problem solved with AI, and 'Expected users' (labeled 1.1) asking for a description of the expected users of the model. Each question has a 'Comment' and 'Attachment' button below it.

2. Go ahead by reading and answering the questions. For the purposes of our exercise, it is more important to understand the nature of the **Questions** themselves, therefore feel free to provide answers such as “ABC” & “No” :

The goal of this questionnaire is to identify AI risks associated with your use case. In this questionnaire, the term "use case" refers to the scenario or application in which AI is used to address a problem. "Model" refers to an AI model that will be used to produce the desired outputs for your use case. You do not need to have a model selected at this point to answer this questionnaire. *

I acknowledge

Clear

Comment Attachment Activity

(1) Problem statement
Please describe the problem you're solving with AI. *

ABC

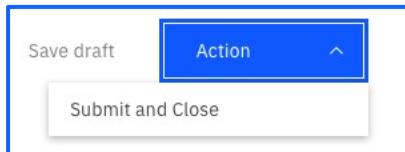
Comment Attachment Activity

(1.1) Expected users
Please describe the expected users of the model. *

The expected users of the model are those who use the model's output.

ABC

- Once all **Questions** are answered, go ahead and select **Submit and Close** under the **Action** button:



- From the top, Click on the tab to navigate back to the **Insurance Claim Summarization Use Case**:

Save draft Action

Submit and Close

Have you already selected a model for implementing this use case?

Yes

No

Clear

Comment Attachment Activity

Note that the **Status** from **Proposed** has progressed to **Awaiting Use Case Approval**.

The screenshot shows the IBM Watsonx Governance console interface. At the top, there's a navigation bar with links like 'Home', 'Insurance cl...', 'Use Cases', '+ New Use Case', and 'AI Risk Ident...'. Below the navigation, the main area displays a 'Use Case' card for 'Insurance claim summarisation'. The card includes sections for 'General', 'Use Case Details', and 'Risk'. In the 'General' section, the 'Status' is set to 'Awaiting Use Case Approval' (highlighted with a red box). The 'Risk Level' is also set to 'Low'. On the right side of the card, there's a sidebar with sections for 'Initial Approval', 'Tags', and 'Select an action to validate'. The 'Initial Approval' section contains a note about reviewing initial details and an 'Actions' button.

5. Additionally, further down in the **Risk** section, we can see one (or many depending on your answers) Use Case Risk identified.

The risk(s) was populated in the Use Case based on our answers on the Questionnaire in the previous step. The risks are configured as a Response Action to the question and come from the AI Risk Atlas, which is an IBM Research produced set of risks associated with Gen AI and Traditional ML. It can be found here: [AI risk atlas - IBM Documentation](#)

The screenshot shows the 'Risk' section in the IBM Watsonx Governance console. It includes two tables: 'Risk Identification Assessments' and 'Use Case Risks'.

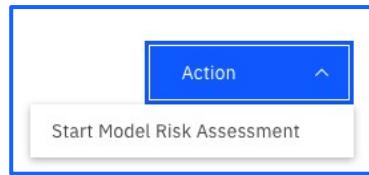
Name	Description	Progress (%)	Tags
AI Risk Identification (Insurance claim summarisation)	AI use case risk identification assessment (2025-01-30)	100	

Name	Description	Inherent Risk Rating	Residual Risk Rating	Status	Tags
Non-disclosure (Copy of MOD_0000000_RIS_0000038)	Content might not be clearly disclosed as AI generated.	Not Determined	Not Determined	Awaiting Assessment	

6. Next we will assess the risk in the context of our Use Case. Open the Use Case Risk by selecting it:

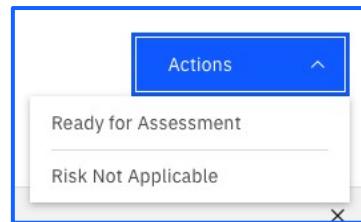
The screenshot shows the IBM Watson Governance console interface. At the top, there are tabs for 'Insurance cl...', 'Use Cases', and 'Non-disclos...'. Below the tabs, the main area displays a risk record titled 'Non-disclosure (Copy of MOD_0000...)' with a status of 'Not Determined' for both Inherent Risk Rating and Residual Risk Rating. The Owner is listed as 'System Administrator'. On the right side, there is an 'Action' button. The left panel contains sections for 'General' (Name: Non-disclosure (Copy of MOD_0000000_RIS_0000038), Status: Awaiting Assessment, Description: Content might not be clearly disclosed as AI generated, Assessment Method: Qualitative), 'Related Content' (Mitigating Controls, Parent Processes, Issues), and 'Risk Context' (Primary Parent, Parent, Child). The right panel includes a 'Tags' section (No tags have been added yet), a 'Risk awaiting evaluation' section (Assess the risk by reviewing the information provided (Controls, Loss Events, KRIs, Processes, Problems). Once completed, submit for risk assessment using the action in the top right corner), and a 'More' link.

7. Go ahead and select **Start Model Risk Assessment** under the **Action** button:



At this stage, the person responsible for the risk assessment (e.g. a Risk Owner) would assess the risk based on controls that exist in the environment. You can see there are options to associate any mitigation controls with the risk itself. This is outside the scope of this lab.

8. Once the **Risk Assessment** is initiated proceed by flagging it as **Ready for Assessment** under the **Action** button:

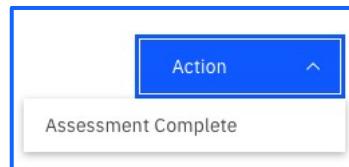


9. Switch to Edit mode and provide **Low** ratings for **Inherent Impact**, **Inherent Likelihood**, **Residual Impact**, **Residual Likelihood**, as well as a placeholder “ABC” for **Migration Strategy**.

The risk level is a subjective rating and calculates an automatic overall risk rating for the specific risk.

The screenshot shows the 'Non-disclosure (Copy of MOD_0000000_RIS_0000038)' risk assessment record in edit mode. The 'Risk Assessment' section contains fields for Inherent Impact (Low), Inherent Likelihood (Low), and Residual Risk Rating (Low). The 'Monitoring & Mitigation' section includes a table for Controls, with columns for Name, Description, Control Owner, Control Type, Operating Effectiveness, Status, and Tags. A modal window titled 'Perform Risk Assessment' is open, showing the stage 'Perform Risk Assessment (Awaiting Assessment)', due date '2/1/2025', and a list of items requiring attention: 'Inherent Impact', 'Inherent Likelihood', 'Residual Impact', 'Residual Likelihood', 'Owner', and 'Description'. The status bar at the bottom indicates 'Assessment Complete'.

- Finally, **Save** and proceed as **Assessment Complete**. Feel free to add a Workflow Comment with this.



- Once the Risk Assessment is completed, navigate back to the Use Case.

You will see the Risk(s) associated with the Use Case and the fact it is “Accepted”. You will also see a series of dashboards and risk ratings which can be configured to show the current risk profile of your Use Case. This is a useful visualisation.

It is now the responsibility of the Model Owner to provide a Risk Rating commensurate with the outcome of the Risk Assessment process. This is subjective based on the Risks visible in the Use Case.

- To do this, scroll down to the **Risk** section, switch to Edit mode and set the Risk Level to **Low**:

The screenshot shows the IBM Watsonx Governance console interface. At the top, there's a navigation bar with 'IBM Watsonx | Governance console' and a search bar containing 'Insurance cl...'. Below the navigation is a breadcrumb trail: 'Insurance claim summarisation' with a star icon and a dropdown arrow.

The main content area has tabs: 'Task' (selected), 'Activity', 'Admin', and 'Security Performance Monitoring'. Under 'Task', there's a section for 'Proposed Solution' and 'Additional Details' with date pickers for 'Data Gathering Completion Date' (mm/dd/yyyy) and 'Risk Assessment Completion Date' (mm/dd/yyyy). There are also 'Search' and 'Copy from Library' buttons.

A 'Risk' section follows, with a dropdown for 'Risk Level' set to 'Low'. It includes fields for 'Risk Identification Completion Date' and 'Risk Assessment Completion Date', both with date pickers. There are 'Add' and 'New' buttons for tags. A progress bar labeled 'Progress (%)' is shown at 100% completion.

The 'Tags' section indicates 'No tags have been added yet.' A detailed view for 'Use Case Data Gathering' is open, showing a stage of 'Use Case Data Gathering (Data gathering)' and a due date of '2/4/2025'. It also lists 'All Key Items (5)' with categories: Purpose (selected), Risk Level, Use Case Type, Stakeholder Departments, and Technical Owner.

At the bottom, there's a table for 'Use Case Risks' with columns: Name, Description, Inherent Risk Rating, Residual Risk Rating, Status, and Tags. A 'Search' bar is also present.

While this risk assessment comes out of the box, it is simple to create your own Risk Assessments based on your own set of organisational risks. This is outside the scope of this lab.

EU AI Act applicability Assessment

The EU AI Act is a regulation affecting AI Use Cases. It categorises Use Cases depending on the purpose of them from Prohibited, High, Medium and Out of Scope. It is important to understand the applicability of your AI Use Case to the regulation, as there are regulatory fines for non-compliance. If the result was prohibited, you would immediately know that this would never be a production use case hence saving yourself time developing this Use Case.

This pre-configured assessment helps an organisation know under which category of the act their AI Use Case falls under.

1. Complete the **Applicability Assessment** under **Regulatory Information**.
2. Once completed, go back to the Use Case view. You will see the associated category of your Use Case.

The screenshot shows the IBM Data Platform interface for the 'Insurance claim summarisation' task. The top navigation bar includes a home icon, a search bar with the text 'Insurance cl...', and a close button. Below the navigation is a breadcrumb trail: 'Insurance claim summarisation' with a star icon and a dropdown arrow.

The main content area has a header 'Task' followed by tabs: Activity, Admin, Security, and Performance Monitoring. A section titled '* Modified Required *' contains four risk rating fields: Fairness Risk Rating ('Not Determined'), Fairness Risk Score, Performance Risk Rating ('Not Determined'), and Performance Risk Score.

A 'Regulatory Information' section follows, containing an EU AI Risk Category ('Not Determined') and an Applicability Assessment Completion Date.

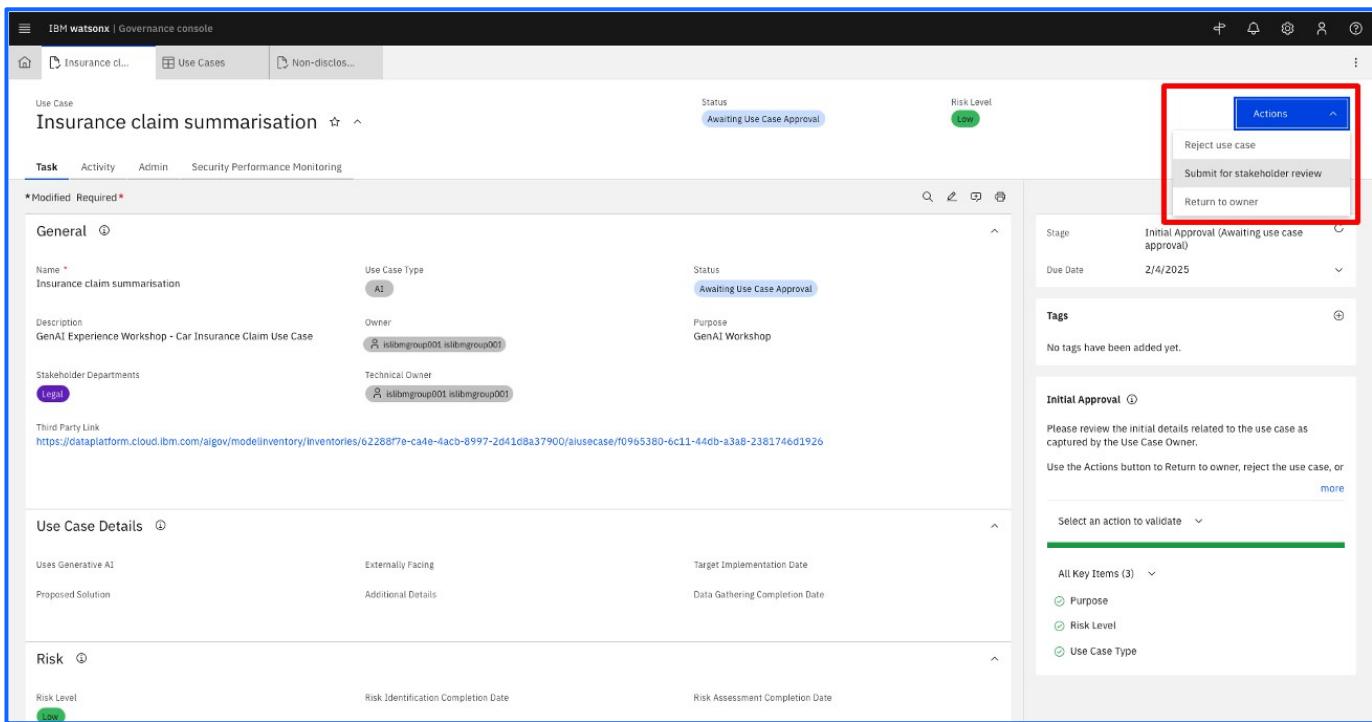
The 'Applicability Assessments' tab is selected, showing a table with columns: Name, Description, Progress (%), and Tags. One row is listed: 'Applicability Assessment (Insurance claim summarisation)' with a description of 'AI use case applicability assessment (2025-01-30)', progress of 0%, and a 'Tags' column with a single tag 'Watsonx Innovation Studio'. There are 'Search' and 'Add' buttons at the top of the table.

Submit for stakeholder review

Assuming this wasn't prohibited, we are now ready to submit this for stakeholder review. The list of stakeholders indicate the different disciplines involved in the sign off of an AI Use Case. This is a final step before the AI Use Case is signed-off for development and ensures that the requisite stakeholders in an organisation have the opportunity to review all the info relating to the Use Case. If they are not happy with this they can reject it or request further information.

For this step, you will be taking on the persona of someone responsible from the legal team signing off on the Use Case.

1. Switch back to the **Use Case** and proceed with **Submit for stakeholder review**:



The screenshot shows the IBM Watsonx Governance console interface. On the left, there's a navigation bar with 'Insurance cl...', 'Use Cases', and 'Non-disclos...'. The main area displays a 'Use Case' card for 'Insurance claim summarisation'. The card includes sections for 'General', 'Use Case Details', and 'Risk'. In the 'General' section, fields like 'Name' (Insurance claim summarisation), 'Use Case Type' (AI), 'Owner' (ibmgroup001), and 'Purpose' (GenAI Workshop) are visible. The 'Status' is listed as 'Awaiting Use Case Approval'. On the right side, there's a 'Actions' dropdown menu with options: 'Reject use case', 'Submit for stakeholder review' (which is highlighted with a red box), and 'Return to owner'. Below the actions, there's a 'Stage' section showing 'Initial Approval (Awaiting use case approval)' and a 'Due Date' of '2/4/2025'. A 'Tags' section indicates 'No tags have been added yet'. At the bottom, there's an 'Initial Approval' section with instructions and a 'Select an action to validate' dropdown. Underneath the validation dropdown, there's a list of 'All Key Items (3)': 'Purpose', 'Risk Level', and 'Use Case Type', each with a green checkmark.

2. Once submitted, it is ready to be reviewed by one of the stakeholders, as it can be seen in the **Use Case Approvals** section:

The screenshot shows the IBM Watson Governance console interface. The main header bar includes links for Home, Use Cases, Non-disclos..., Task, Activity, Admin, and Security Performance Monitoring. The current page is 'Insurance claim summarisation'.

Use Case Approvals

Stakeholder Approval Completion Date: [empty]

Use Case Reviews:

Name	Stakeholder Departments	Approval Date	Approval Status	Risk Rating	Review Comment	Tags
Insurance claim summarisation-Review-00003 Watson Innovation Studio	Legal		Awaiting Approval	Not Determined		

Related Models

All Models | Prompts and Tunes | Associated Foundation Models | Deployments | Model Links | Model Groups

Name	Description	Model Owner	Model Class	Model Status	Tags
No results					

Other Relationships

Stakeholder Review

No tags have been added yet.

Any relevant stakeholder departments that were selected in the data gathering stage have been sent a **Use Case Review** task to obtain their approval for this use case. Once all of the approvals are received, this use case can be approved for development.

Approve for development

1 item requires attention.

All Key Items (4)

- Purpose
- Risk Level
- Use Case Type

Key Items for this Action

Logic: 1 OR 2

- 1: Use Case Reviews
- 2: Use Case Reviews

3. Open the **Use Case Review** by selecting it:

Name	Stakeholder Departments	Approval Date	Approval Status	Risk Rating	Review Comment	Tags
Insurance claim summarisation-Review-00003	Legal		Awaiting Approval	Not Determined		

4. Set the **Reviewer** as your current username, for **Review Comment** type “Approved”, set the **Risk Rating** as Low and press **Save**:

5. Once Saved, **Approve Use Case** under the Action Button:



6. Return to the Use Case. You can now **Approve for Development**:



7. The Use Case has progressed through the Plan phase and had the necessary approvals from key stakeholders including Risk Management and Legal. This Use Case is now “Approved for Development”.

The screenshot shows the 'Insurance claim summarisation' Use Case in the IBM Watson Governance console. The 'Status' field is highlighted with a red box and displays 'Approved for Development'. The 'Risk Level' is indicated as 'Low' with a green icon. The 'General' section includes fields for Name (Insurance claim summarisation), Use Case Type (AI), Owner (ibmgroup001), and Purpose (GenAI Workshop). Stakeholder Departments include Legal. A Third Party Link is provided. The 'Use Case Details' section shows Uses Generative AI, Proposed Solution, Externally Facing, Target Implementation Date, Additional Details, and Data Gathering Completion Date. The 'Risk' section shows Risk Level as Low, Risk Identification Completion Date as 1/30/2025, and Risk Assessment Completion Date as 1/30/2025. A sidebar on the right provides a general view of the Use Case and lists key items: Purpose, Risk Level, and Use Case Type.

8. Scroll down to the bottom of the Use Case to see that there has been a relationship diagram developed that ties the Use Case to the various objects recently developed.

The screenshot shows the 'Relationships' section for the 'Insurance claim summarisation' Use Case. A relationship diagram is displayed, showing the Use Case connected to four entities: Business Entities, Model Groups, Questionnaire Assess., and Risks. The diagram uses blue circles for Primary Parent and Parent nodes, and black circles for Child nodes. The 'Relationships' section also includes a legend: Primary Parent (blue circle), Parent (blue circle), and Child (black circle). The sidebar on the right remains the same as in the previous screenshot.

9.

Well done, you have successfully completed Phase 1 of the AI Use Case lifecycle, the Planning phase. You have identified:

- the purpose

- technical owners
- the Risk Level
- whether it is compliant to the EU AI Act - stakeholder approval.

Please pause here before we explain the next Phase.

Phase 2: Design

Create a Prompt Template

Let's begin creating our first prompt using Watsonx.AI. Watsonx.AI is IBM's AI studio that allows you to build, tune and deploy AI models.

You can click on the third party link within the Use Case to navigate to the Watsonx.AI platform which will show you the Use Case. Separately use this link: <https://ibm.biz/BdGSi2>

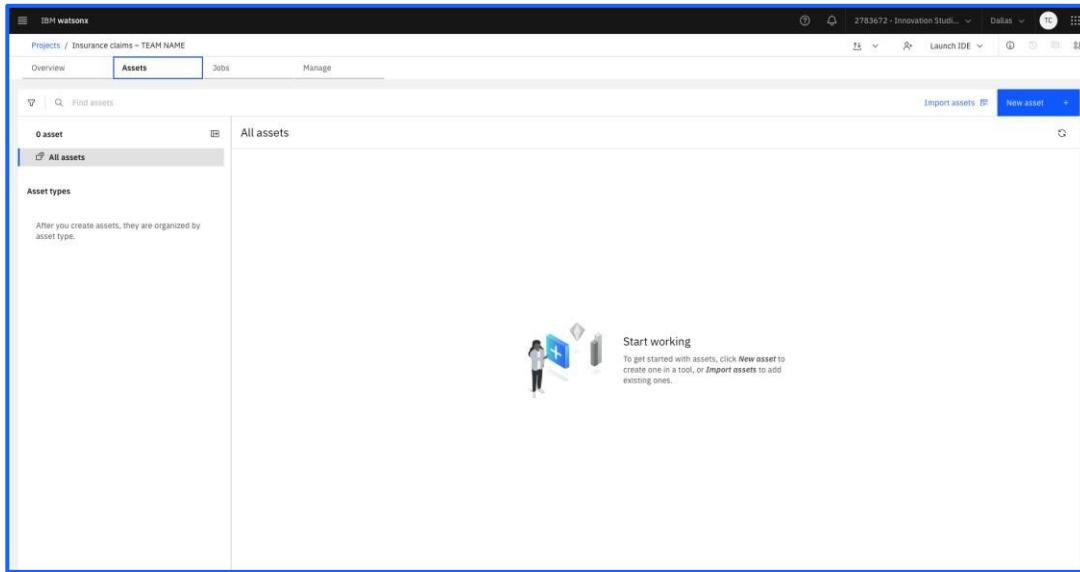
The screenshot shows the IBM Watsonx Governance console interface. A specific use case entry is selected: "Insurance claim summarisation - sammclear". The "Third Party Link" field contains the URL <https://dataplatform.cloud.ibm.com/aigov/modelinventory/inventories/62288f7e-ca4e-4acb-8997-2d41d8a37900/aiusecase/ae86c7d7-11>.

From IBM watsonx.AI, use the hamburger menu to navigate to your Team's project e.g. Team 1.

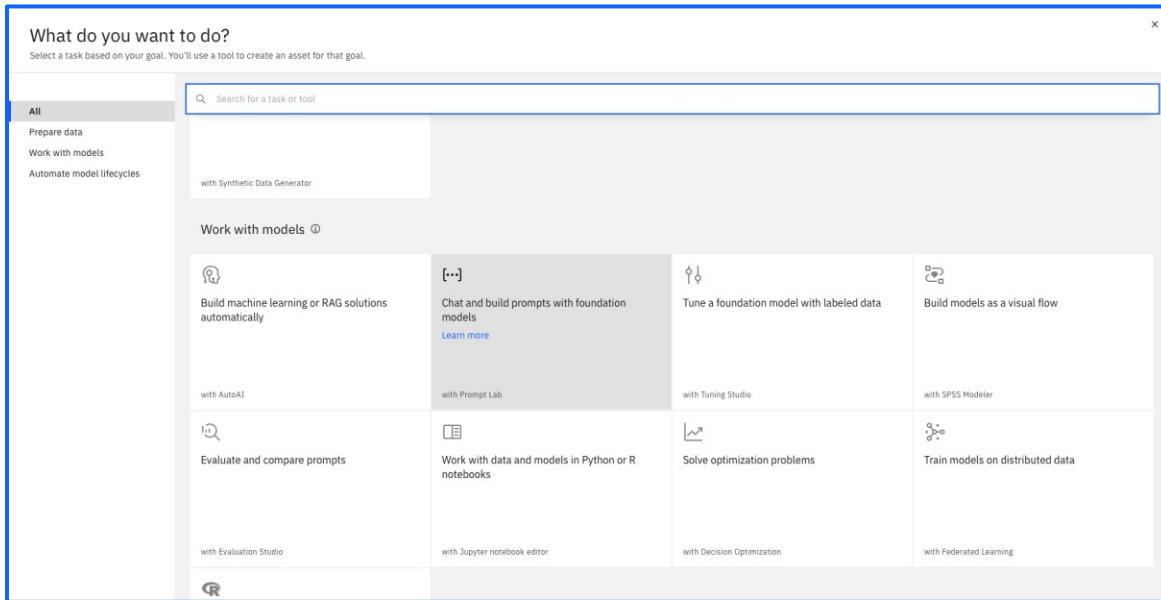
The screenshot shows the IBM Watsonx AI use case details page for "Insurance claim summarisation - sammclear". The "Status" is listed as "Approved (System, Feb 11, 2025)". The "Risk level" is set to "Low". The "Inventory" section lists "WatsonX Workshop". The "Tags" section indicates "Add tags to this AI use case.".

Projects are spaces to collaborate with others and build data and AI assets. You will see multiple options along the top which include information about the project, the assets associated with the project, any current jobs that are running and to allow the management of that asset.

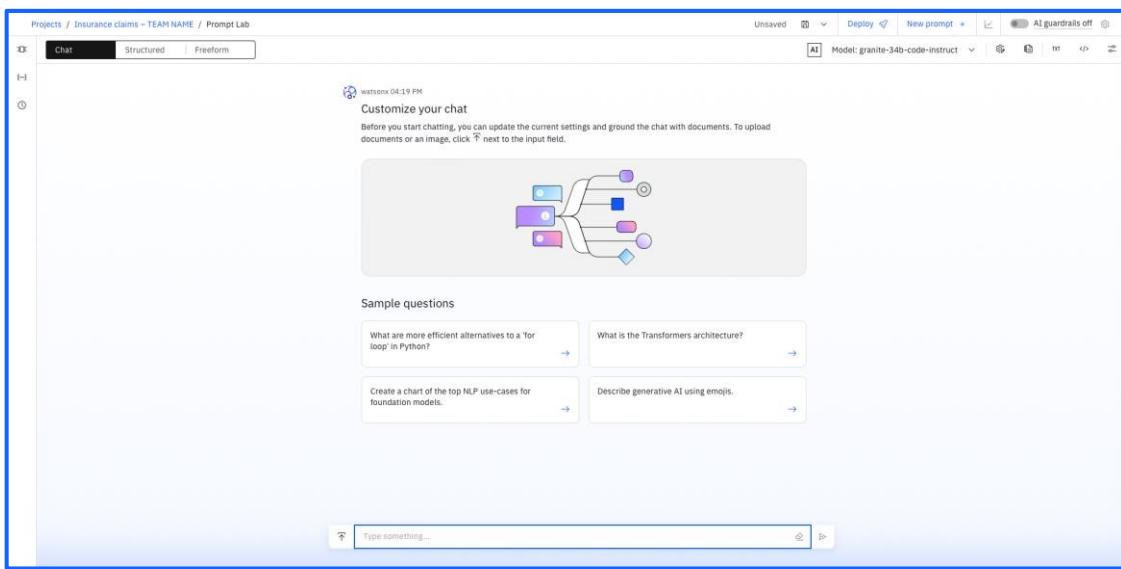
Head over to the **Assets** tab of the project home page. Once there, click on **New Asset**. Note that your project will already include assets for a later part of the lab.



Now, if you scroll down on the menu that appears, you should see a card that says **Chat and build prompts with foundation models**.



You will be taken into the Watsonx.AI prompt lab. If there is an option to take a tour, skip it. You will be met with the prompt lab. This screen has numerous options on how you can configure your prompt to suit your needs.

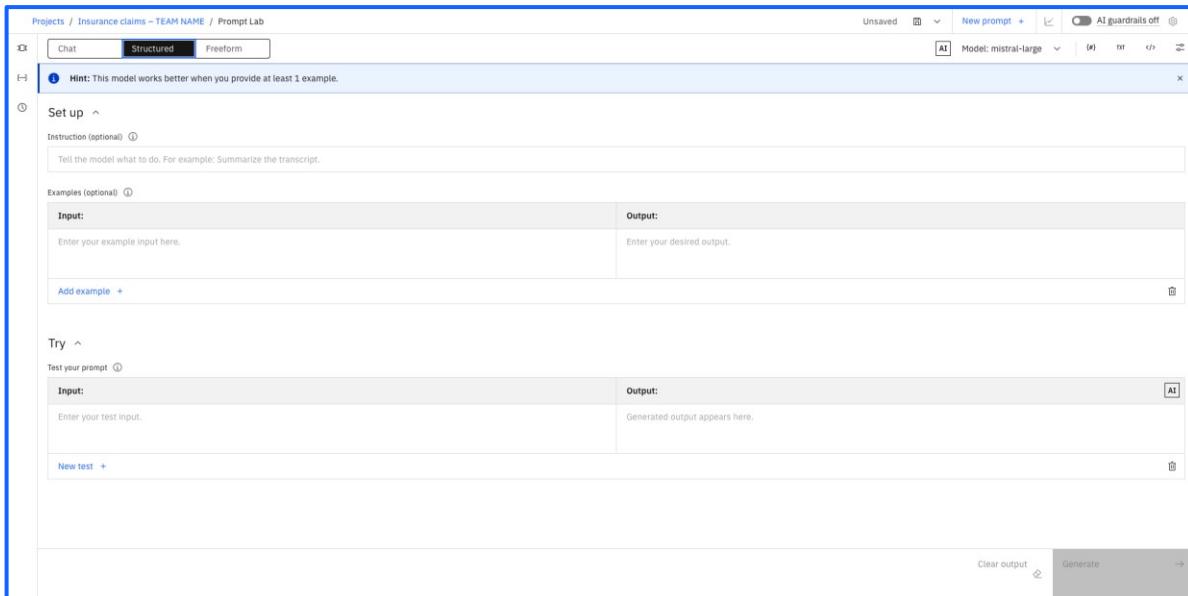


Right above that, you see the option to select “Chat”, “**Structured**” or “Freeform”.

Structured prompts have easily identifiable sections to place examples, instructions, and prompt variables (which we will discuss later). Freeform prompts have a looser and more flexible structure. With Chat mode, you can chat with the foundation model to see how the model handles dialog or question-answering tasks.

This lab will walk through utilising the **Structured** prompt functionality, however this is possible to do with freeform as well, so choose whichever option is the most comfortable for you!

You will notice that there is a box labelled **Instruction** which is where we will be placing the bulk of our prompt to be evaluated later.



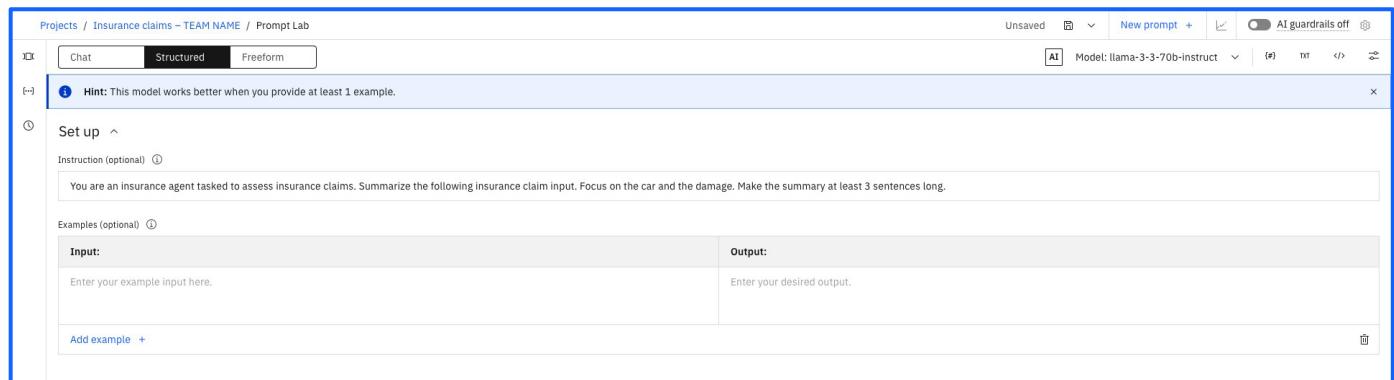
You will also see an icon with a graph at the top right , which is the button to run an evaluation that we will discuss later. You will be able to see the status of your work at the top right as well, indicating whether your prompt session work is Saved or Unsaved.

Select the “llama-3-3-70b-instruct” model from the menu on the top right.

Inside of the Instruction box, enter your first prompt that we will be evaluating throughout this lab. For now, enter the prompt shown in the screen below. This is the instruction given to the model.

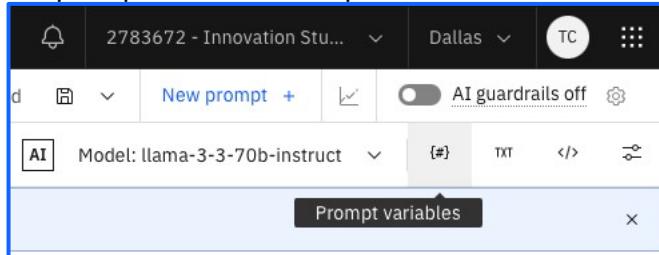
Copy & Paste from in between the quotation marks:

“You are an insurance agent tasked to assess insurance claims. Summarise the following insurance claim input. Focus on the car and the damage. Make the summary at least 3 sentences long.”

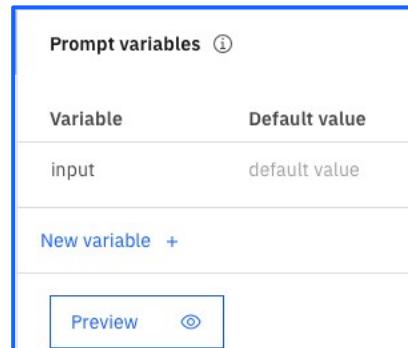


Prompt Variables

Now, we need to ensure that we are able to test our prompt against validation and test data. To do so, we need to ensure we have prompt variables set up for our instructions.



Once you have opened up this sidebar, you will see two empty boxes pop up, with **Variable** and **Default Value** able to be filled out. Use "input" as the variable name and leave the default value empty.



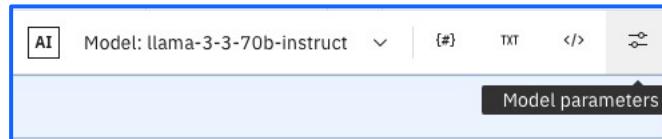
By scrolling down your page, you will notice an area on the screen titled **Try**. Inside of this, you want to put a variable titled "{input}". This variable will be replaced by input insurance claims.

A screenshot of the 'Prompt Lab' interface. On the left, there are tabs for 'Chat', 'Structured', and 'Freeform'. The 'Structured' tab is selected. In the center, there's a 'Set up' section with an optional instruction and examples. On the right, the 'Prompt variables' sidebar is visible, showing the 'input' variable with a 'default value'. Below the sidebar is a 'Preview' button. In the bottom right corner of the main area, there's a 'Try' section. Inside the 'Try' section, there's an 'Input' field containing the placeholder '{input}' (which is highlighted with a red box). To the right of the input field is an 'Output' field with the placeholder 'Generated output appears here.' At the very bottom right of the interface are 'Clear output', 'Generate', and a right-pointing arrow buttons.

Model Parameters

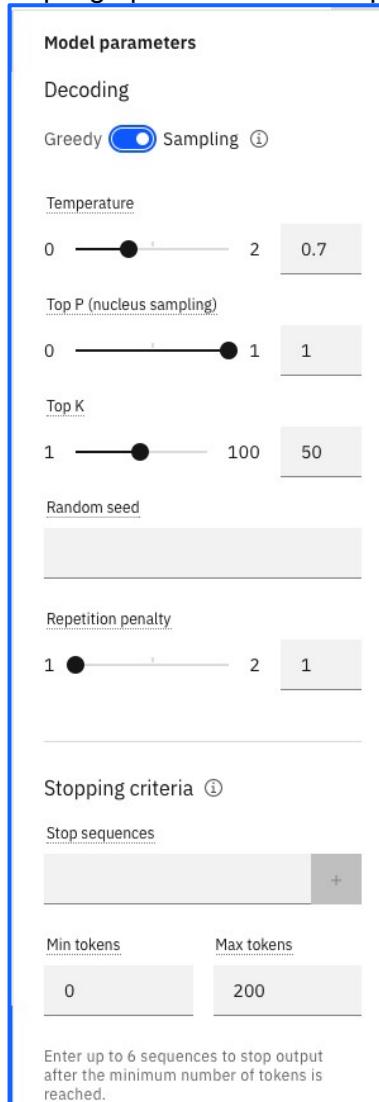
Next to where you clicked prompt variables, you should see an option that says **Model parameters**. Click on that next. We will not be making any serious changes here just yet but mostly giving a tour of this section so you can make modifications later.

Most importantly, when you select the button highlighted below (top right), you will see a toggle for **Greedy** or **Sampling**.



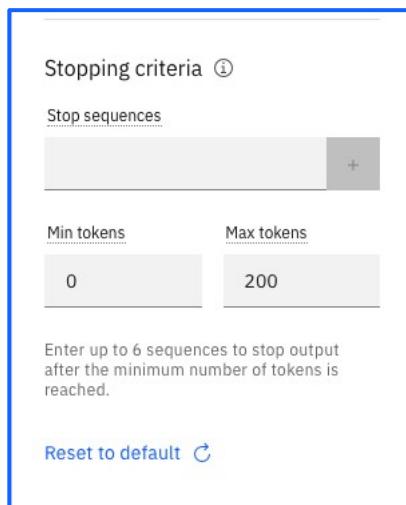
Selecting greedy will allow your model results to be reproduceable across experiments, while selection sampling will give you more variability.

For our task, we will be choosing the sampling option for this base prompt.



Now that you have chosen sampling, you will see several new options pop up that are specific to the sampling decoding method. **Temperature** is the most important factor here, where setting a higher temperature will give higher variation and more unique outputs. Hover over the info icons to gain more information on **Top P** and **Top K** methods as well.

Lastly, scroll all the way down the **Model parameters** sidebar to find the **Stopping criteria** as well as **Token limits**. For LLMs, tokens ~ words, so we want to set that higher (e.g. 200) to get more detail in our summaries of the insurance claims.



Saving Prompt

Once you are all finished, click the **Save** icon dropdown in the top right, and then click **Save as**. This will take you to the following screen, where you will select your asset type, prompt template name, and task.

The screenshot shows the 'Prompt Lab' interface within the 'Insurance claims - TEAM NAME' project. In the 'Set up' section, there is an 'Instruction (optional)' field containing the text: 'You are an insurance agent tasked to assess insurance claims. Summarize the following insurance claim input. Focus on the car and the damage. Make the summary at least 3 sentences long.' Below it, an 'Examples (optional)' section has a table with one row. The 'Input' column contains 'Enter your example input here.', and the 'Output' column contains 'Enter your desired output.'. A blue button labeled 'Add example +' is at the bottom. In the 'Try' section, there is a 'Test your prompt' field with an 'Input' column containing '{input}' and an 'Output' column containing 'Generated output appears here.' A blue button labeled 'New test +' is at the bottom. At the bottom right of the interface are 'Clear output' and 'Generate' buttons.

Make sure to select the **Prompt template** option, then fill out a name in a similar fashion shown in the picture. From there, select **Summarization** for your task, and make sure to select **View in project after saving**.

The screenshot shows the 'Save your work' dialog box. Under 'Asset type', the 'Prompt template' option is selected, with the note: 'Save the current prompt only, without its history.' Other options are 'Prompt session' (Save history and data from the current session) and 'Standard notebook' (Save the current prompt as a notebook). On the right, under 'Define details', the 'Name' field is set to 'Insurance Claim Summary Prompt'. The 'Task' dropdown is set to 'Summarization'. There is also a 'Description (optional)' field and a checkbox for 'View in project after saving'. At the bottom are 'Cancel' and 'Save' buttons.

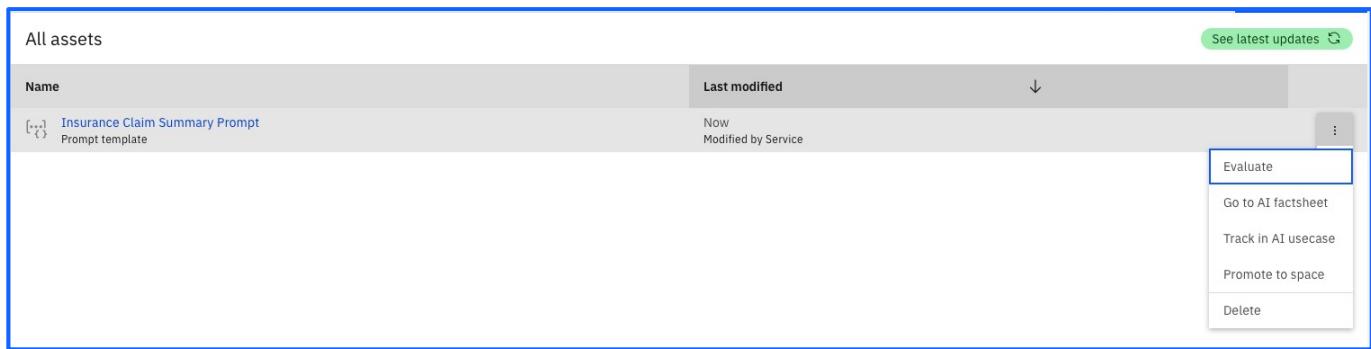
You have now successfully designed your first prompt template.

Phase 3: Development

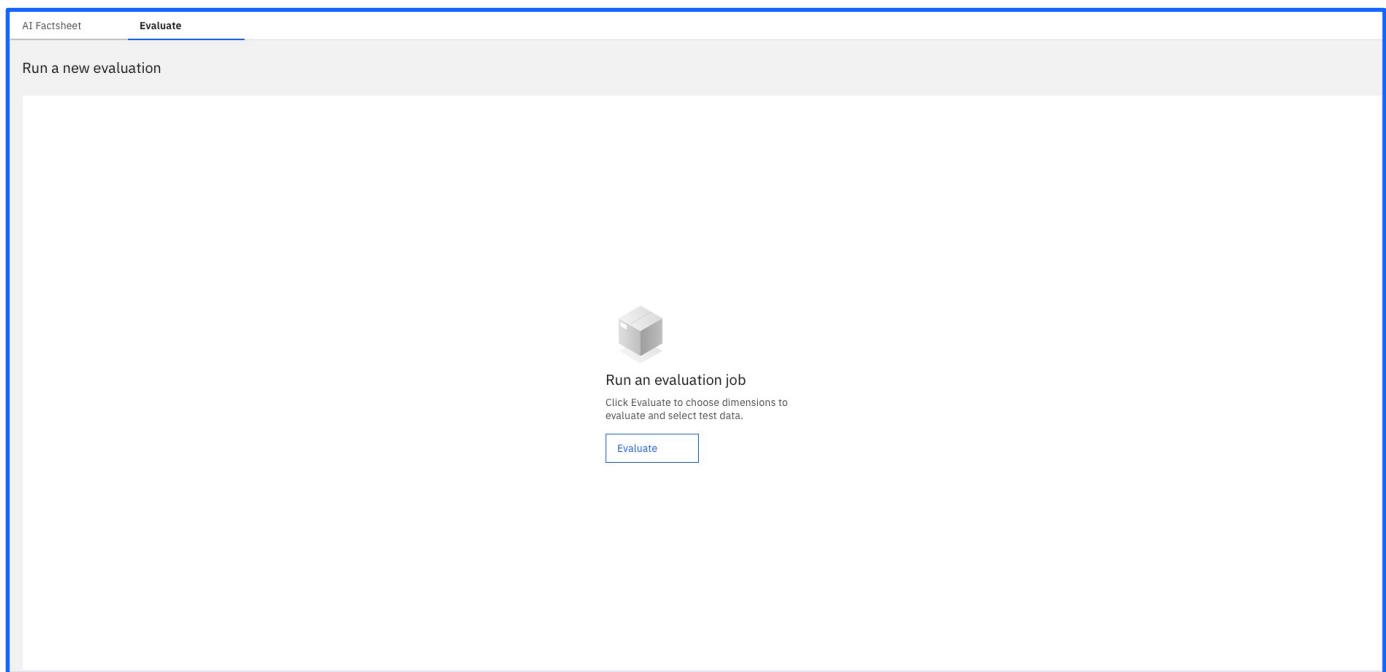
Evaluating your prompt template

The goal of this part of the lab is to iterate on our first prompt that we have created, and work to achieve better results so that we can deploy our prompt and eventually bring it to production.

Now you should see your finished prompt template on the screen when you are taken back to the **Assets** page. Verify that you have completed the previous steps correctly by clicking the 3 dots to the right of the prompt and choose **Evaluate**. We are going to be testing out our new prompt!



You should now see a screen pop up with an evaluate button in the middle. Select that button.

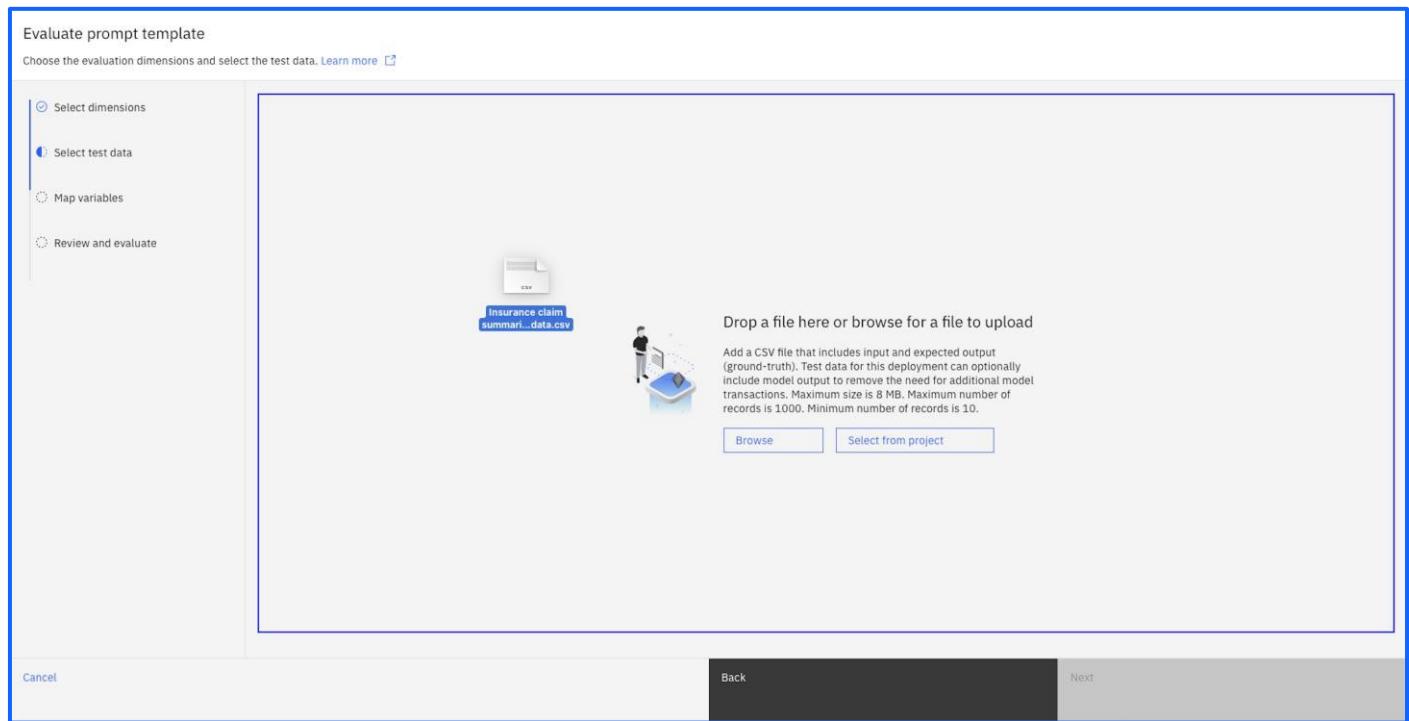


Now you will see a screen asking you which dimensions to evaluate. At this point, we are only able to evaluate GenAI quality (model health is enabled by default). By clicking on the arrow next to the Generative AI Quality, you can see which metrics are being calculated and what the definition of those are. It is also possible to switch metrics on/off via the Advanced settings feature.

A screenshot of the 'Evaluate prompt template' wizard, currently on the 'Select dimensions' step. On the left, a sidebar lists four steps: 'Select dimensions' (selected), 'Select test data', 'Map variables', and 'Review and evaluate'. The main panel shows a title 'Select dimensions to evaluate' with a sub-instruction: 'These dimensions are based on the prompt template task type.' Below this is a table with two columns: 'Dimension' and 'Description'. The table contains two rows: 'Generative AI Quality' (with a checked checkbox) and 'Model health' (with an unchecked checkbox). A link 'Advanced settings' is located at the top right of the table. At the bottom of the panel are 'Cancel', 'Back', and 'Next' buttons. The 'Next' button is highlighted in blue.

Now it's time to choose our test data. For this tutorial, we will be using the **Insurance claim summarisation validation data.csv**

Choose Select from Project and locate the file.



Now it's time to map the input variables to the input variable that we created earlier. Additionally, I chose **Input** to be mapped to the “Insurance_Claim” column in my file. This will ensure that my prompt is focused on generating the correct data.

Lastly, **Reference output** is the column of the data that gives the expected output from our LLM where metrics will be evaluated against. This is how we can measure how effectively we have structured our prompts.

I chose the reference output to be mapped to ‘Summary’ – this is the column of data that gives the expected output to the LLM.

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

Select dimensions

Select test data

Map variables

Review and evaluate

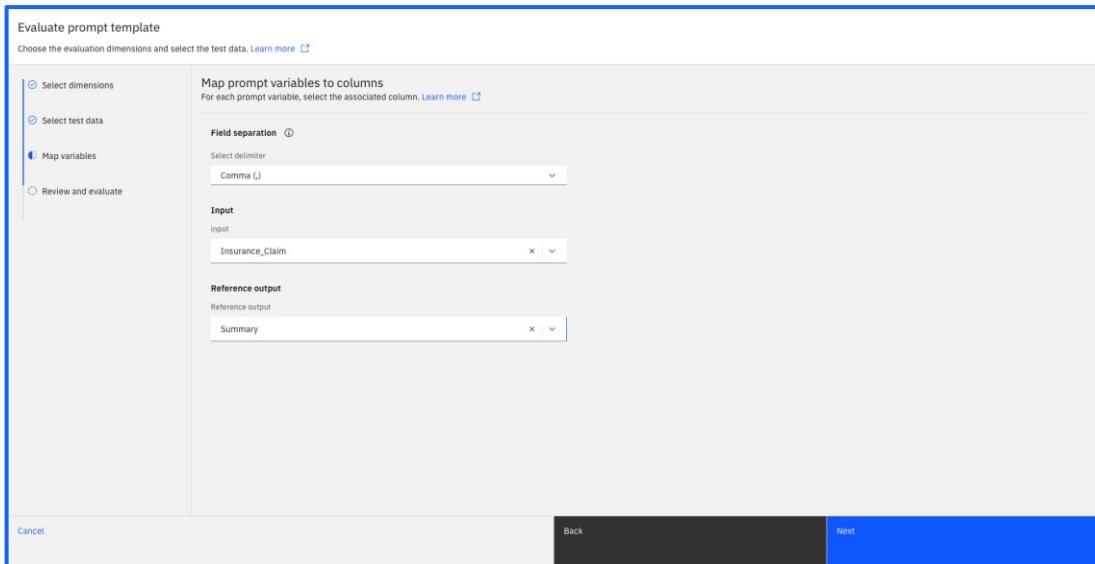
Map prompt variables to columns
For each prompt variable, select the associated column. [Learn more](#)

Field separation [?](#)
Select delimiter: Comma (,)

Input
input: Insurance_Claim

Reference output
Reference output: Summary

[Cancel](#) [Back](#) [Next](#)



After that, you will see a screen to review the information you have entered. Once you verify that you have entered everything correctly, hit the **Evaluate** button. and wait until your prompt has finished being evaluated. This process can take up to a few minutes.

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

Select dimensions

Select test data

Map variables

Review and evaluate

Review

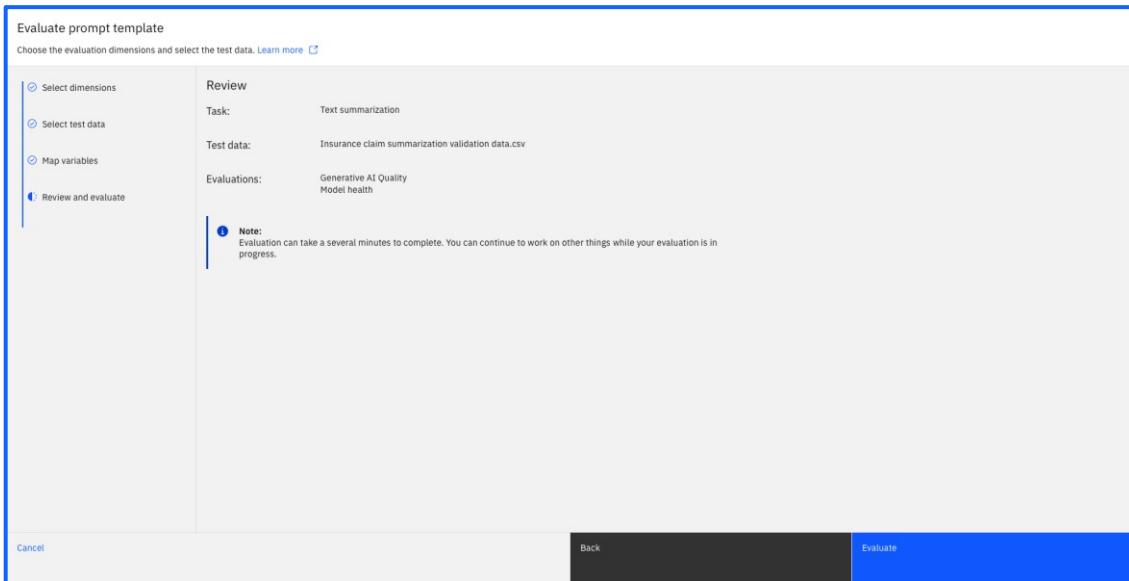
Task: Text summarization

Test data: Insurance claim summarization validation data.csv

Evaluations: Generative AI Quality
Model health

Note:
Evaluation can take a few minutes to complete. You can continue to work on other things while your evaluation is in progress.

[Cancel](#) [Back](#) [Evaluate](#)



Evaluation Results

Once the evaluation has finished, you should see a screen pop up displaying the results. As you can see here, our evaluation was not very successful, triggering a number of alerts and failing its test. This is to be expected!

The screenshot shows the AI Factsheet Evaluate tab for a project named "Insurance claims – TEAM NAME / Insurance Claim Summary Prompt". It displays the following information:

- Deployment details:** Test data set is "Insurance claim summarization validation data.csv".
- Test details:** 1 test run, 0 tests passed, 1 test failed.
- Model health:** 10 records, 8,727 ms median record latency, 2,392 total input token count, 2,000 total output token count.
- Generative AI Quality - Text summarization:** Alerts triggered: 15. A table shows scores for various metrics:

Metric	Score	Violation
SARI	42.57	37.43
METEOR	0.39	0.41

Click on the blue arrow → on the right of **Generative AI Quality – Text Summarisation** to access a more detailed view of the evaluation results. Make sure to check both the **Model health** and **Generative AI Quality** tabs.

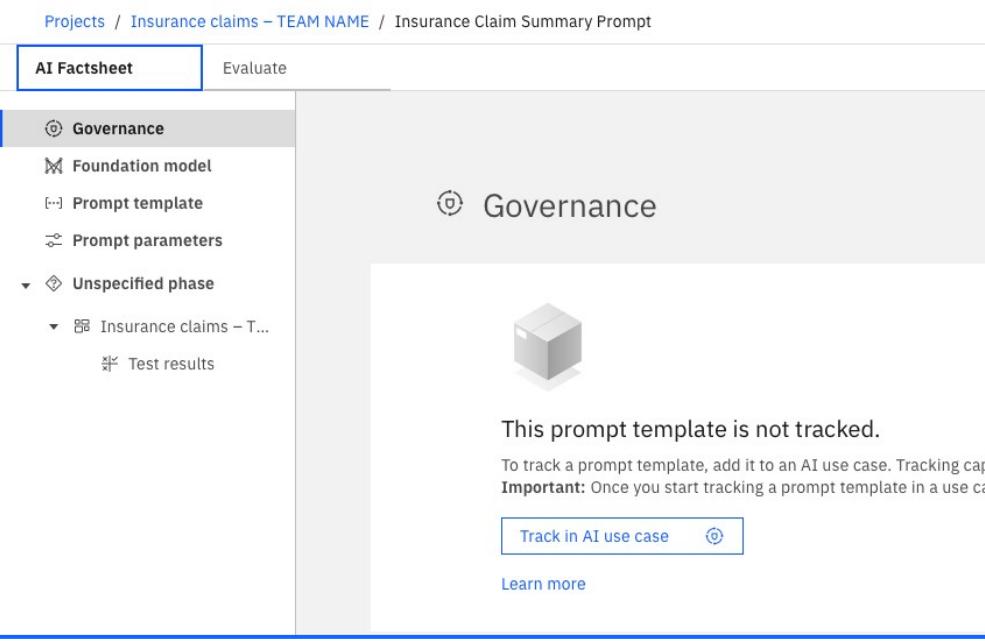
The screenshot shows the Generative AI Quality tab for the same project. It displays the following information:

- Dataset:** Feedback data.
- Metrics:** Coverage (0.63), ROUGE-1 (0.23), SARI (42.6), METEOR (0.39), F1 Score (0.23), BLEU (0.06), Jaccard similarity (0.19), and Readability (45.4).
- Content Analysis:** A collection of metrics used to evaluate the model output against the model input or context. It includes Coverage (0.63), Density (0.01), Compression (0.98), and Repetitiveness (0.15).
- Chart:** A scatter plot comparing Coverage (Y-axis) against another metric (X-axis). The X-axis has points at 0.98, 0.50, 0.25, and 0.12. The Y-axis has points at 0.98, 0.50, 0.25, and 0.12. Data points are colored red, green, blue, and purple.

In this lab's use case, the LLM's task was “Summarisation” which can be measured using a range of quality metrics. We recommend taking time now to scan through this documentation on the [generative AI quality metrics supported by watsonx.governance](#).

AI Factsheet

Up in the top left corner of your screen, choose the **AI Factsheet** tab, which will show us a better breakdown of the results we got. The first thing you will see on the page is an option to track your prompt in an AI use case.



The screenshot shows the 'AI Factsheet' tab selected in a navigation bar. The main content area is titled 'Governance'. On the left, a sidebar lists categories: 'Foundation model', 'Prompt template', 'Prompt parameters', 'Unspecified phase', and 'Test results'. Under 'Unspecified phase', 'Insurance claims – T...' is expanded. A central panel displays a 3D cube icon with the text 'This prompt template is not tracked.' Below it, a note says 'To track a prompt template, add it to an AI use case. Tracking cap...'. An 'Important' note states 'Once you start tracking a prompt template in a use ca...'. A 'Track in AI use case' button is at the bottom, along with a 'Learn more' link.

First explore the page by clicking into each of the tabs:

- The *Governance* tab collects basic information such as the name of the AI use case, the description, and the approach name and version data.
- The *Foundation model* tab displays the name of the foundation model, the license ID, and the model publisher.
- The *Prompt template* shows the prompt name, ID, prompt input, and variables, whilst the *Prompt parameters* collect the configuration options for the prompt template, including the decoding method and stopping criteria.
- Finally, the *Evaluation / Unspecified phase* tab displays the data from evaluation, including alerts and metric data from the evaluation. Within the *Test* sub-tab you'll see a screen similar to the one shown below:



Note: The metrics that you see in the lab may differ slightly to what you see above.

You can see that each metric has an upper and lower bounds with the ideal range indicated by the bolder portion at the left side of each metric's measurement. The upside-down triangles represent the measured value for each metrics.

Now, I'll show you how to modify the thresholds for each metric so that your prompt can have looser requirements to pass testing. In a real-world use case, this should only be done if necessary and makes sense for your use case! Head back over to the evaluation screen, where you will see a blue settings button to the right of **Generative AI Quality – Text Summarisation** where we can

adjust our thresholds.



Last evaluation: Tue, Jan 28, 2025 at 12:58 PM GMT

Test details

Tests run	1
Tests passed	0
Tests failed	1

Model health

Records	10
Latency (record)	8,727 ms Median record latency
Token count	2,392 Total input token count
	2,000 Total output token count

Generative AI Quality - Text summarization

Alerts triggered: 15

Metric	Score	Violation
Content Analysis		
ROUGE		
SARI	42.57	37.43
METEOR	0.39	0.41
Text quality		

Go ahead and click on that button, and you will be greeted with the following screen:

Insurance Claim Summary Prompt

Model info

- Model details
- Evaluations
 - Fairness
 - Quality
 - Drift v2
 - Explainability
 - Generative AI Quality** (selected)
 - Model health

Generative AI Quality

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

Settings

Task type: Text summarization
Minimum sample size: 10
Selected metrics: Content Analysis, ROUGE, SARI, METEOR, Text quality, BLEU, Sentence similarity, Readability

Content Analysis

Thresholds: Coverage: 0.7, Repetitiveness: 0.2, Abstractness: 0.2

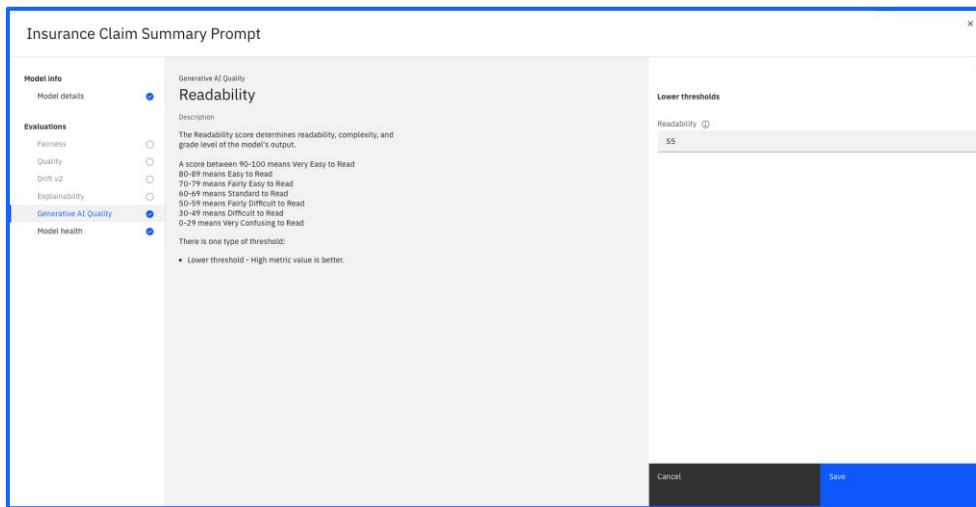
ROUGE

Lower thresholds: ROUGE-1: 0.8, ROUGE-2: 0.8

Here we are going to adjust the **Readability** metric. This metric is higher if your model's generated responses to your prompt are easy to interpret, and lower if they are confusing to read. Scroll down the page and click the edit button to modify it.



Lastly, go ahead and change the readability slightly, say to 55, which puts the requirement into the "Fairly difficult to read" range. After, that, you can go ahead and click save to finalise your changes. Just a reminder here that these metrics should not be altered too much in a real-world use case!



Govern your prompt template

The next section of this lab is to govern your prompt template.

In order to do this, you will need to first associate the project that you created earlier (most likely the Sandbox environment, unless you created one) to the relevant AI Use Case.

To do so, click into the [Track in AI use case](#) button within the **Governance** tab:

The screenshot shows the AI Factsheet interface for a project titled "Insurance claims - TEAM NAME / Insurance Claim Summary Prompt". The left sidebar lists sections: AI Factsheet (selected), Evaluate, Governance, Foundation model, Prompt template, Prompt parameters, Unspecified phase (expanded), Insurance claims - T... (selected), and Test results. The main content area is titled "Governance" and contains a message: "This prompt template is not tracked. To track a prompt template, add it to an AI use case. Tracking captures details about the asset for governance purposes. Important: Once you start tracking a prompt template in a use case, you can no longer edit it. Wait until the prompt template is stable to start tracking." Below this is a "Track in AI use case" button and a "Learn more" link. Further down, under "Foundation model", there is a card for "llama-3-3-70b-instruct" by Meta. At the bottom, there is a "[...] Prompt template" section.

Then, select [Go to AI use cases](#) to open up a list of AI use cases.

The screenshot shows the same AI Factsheet interface as above, but with a modal dialog box in the foreground. The dialog title is "Your project isn't associated to an AI use case yet". The message inside says: "In order to track this AI asset, the project needs to be associated to an AI use case. This ensures that assets within one project are only governed by one AI use case." At the bottom of the dialog are "Cancel" and "Go to AI use cases" buttons, with "Go to AI use cases" being highlighted in blue.

Select the Use Case that you created earlier, and you will be met with this screen:

The screenshot shows the AI use cases interface with the 'Overview' tab selected. Key details include:

- Name:** Insurance claim summarisation - TEAM NAME
- Status:** Proposed (System, Jan 27, 2025)
- Risk level:** None
- Owner:** Murali Krishnan Sankaran (murali.sankaran@ibm.com)
- Tags:** None
- Associated workspaces:** WatsonX Workshop

Scroll down to the **Associated workspaces** section. In this lab, the relevant phase is the **'Develop'**; the prompt requires engineering to develop and further refine it. Click the button shown below to **Associate workspace +**.

The screenshot shows the 'Associated workspaces' section with the following phases:

- Develop:** In this phase your development team will experiment with different approaches of ML development or prompt engineering and test them in projects.
- Associate workspace +** (button highlighted with a red box)
- Validate:**

You will see the following screen and will need to select the relevant **Project** (most likely your Sandbox environment) by ticking the box on the right.

Associate workspaces Develop

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

How it works

Association rules

Associate projects and spaces with phases of your AI lifecycle. This will ensure that assets in these workspaces will be only governed under this AI use case. Multiple workspaces can be associated to one AI use case.

[Read more](#)

Why is this important?

Associate projects and spaces with phases of your AI lifecycle. This will ensure that assets in these workspaces will be only governed under this AI use case. Multiple workspaces can be associated to one AI use case.

[Read more](#)

Projects

A project can only be associated to one lifecycle phase in on AI use case.

Find projects New project +

Name	Created	Your role	Tracked AI assets	Associates

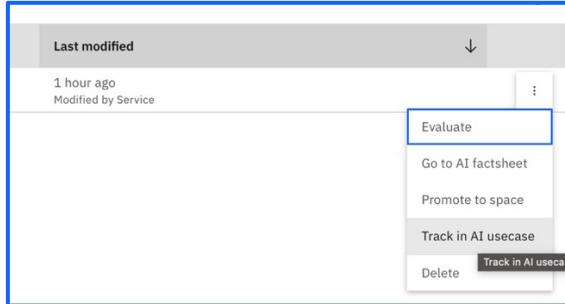
Cancel Save

Tracking Prompt Template in a Use Case

Now that you have associated your project to your AI use case, you can track your prompt template in your AI use case. The idea behind tracking prompt template in a Use Case is to save AI Engineers time and give transparency into the work they're doing. You can automatically save prompt templates and use versioning to easily identify the different prompts and where they are in the lifecycle.

Select the navigation menu in the top right again, and under **Projects**, click into **View all projects** to navigate to your project. Click into this and find the prompt you made earlier. If you struggle to find your way to your project, your facilitator will be able to help.

If you click the three dots on the right, you will be able to see that the prompt needs to be tracked – click on the Track in AI usecase button.



Select the **Default approach**, select to automatically create a **New asset record** and make it **Experimental**.

Insurance Claim Summary Prompt

Track in AI use case

Track an asset to collect details about the asset in factsheets as part of a governance strategy.

Define approach

Define asset record

Assign version

Review

Define approach

Use case: **Insurance claim summarisation – TEAM NAME**

An approach defines one path for solving the goal of the use case. Ex. an approach might be a variation on a machine learning model, or a challenger model. Each approach can include multiple versions.

New approach +



Default approach
A default approach for tracking your AI assets.

[Cancel](#) [Back](#) [Next](#)

Insurance Claim Summary Prompt

Track in AI use case

Track an asset to collect details about the asset in factsheets as part of a governance strategy.

Define approach
 Define asset record
 Assign version
 Review

Define asset record

Approach: **Default approach** | Use case: **Insurance claim summarisation – TEAM NAME**

Associate the trained model with an existing model record or create a new model record in OpenPages. This will sync the tracked model facts between Model inventory and OpenPages.

Asset Record

Existing asset record + New asset record



Create record automatically

A new asset record will be created in OpenPages when model tracking is initiated.

[Cancel](#) [Back](#) [Next](#)

Insurance Claim Summary Prompt

Track in AI use case

Track an asset to collect details about the asset in factsheets as part of a governance strategy.

Define approach
 Define asset record
 Assign version
 Review

Assign version

Approach: **Default approach** | Use case: **Insurance claim summarisation – TEAM NAME**

Choose the starting point for this approach.

 Experimental
Use this as a starting point if your model is just starting in development and its input and output structure will likely change in the near future.
0.0.1

 Stable
Use this as a starting point if your model is in a production state and you won't expect any major changes in its input and output structure soon.
1.0.0

 Custom
Define your own starting version if you already tracked this model in a versioning context before.

Version number
0.0.1

Comment (optional)

Provide context for this change. Comments are saved in the model submit to govern.

[Cancel](#) [Back](#) [Next](#)

Review your choices and start Tracking the Asset

Insurance Claim Summary Prompt

Track in AI use case

Track an asset to collect details about the asset in factsheets as part of a governance strategy.

Define approach
 Define asset record
 Assign version
 Review

Review

Important
 Make sure your prompt template is stable before you track it in an AI use case. Tracking a prompt template locks it so it is no longer editable. Complete edits and evaluations before enabling tracking. You can untrack the prompt template to edit it, but you will lose the tracking history.

Insurance claim summarisation – TEAM NAME

Approach Version
 Default approach ◇ 0.0.1
 A default approach for tracking your AI assets.
 00000000-0000-0000-0000-000000000000

[Cancel](#) [Back](#) [Track asset](#)

Projects / Insurance claims – TEAM NAME / Insurance Claim Summary Prompt [Open in Prompt Lab](#) [Export report](#)

AI Factsheet Evaluate

Governance
 Foundation model
 Prompt template
 Prompt parameters
 Development
 Insurance claims – T...
 Test results
 Additional details
 Attachments

Governance

Insurance claim summarisation – TEAM NAME
 Proposed System | WatsonX Workshop | None | c2428369-0ef4-4f06-996d-04b345363efc

Description
 GenAI Experience Workshop - Car Insurance Claim Use Case

Approach Version
 Default approach ◇ 0.0.1
 A default approach for tracking your AI assets.
 00000000-0000-0000-0000-000000000000

Lifecycle
 01 Develop 02 Validate 03 Operate

[Untrack](#)

Foundation model

Using the hamburger menu, navigate to your **AI Use Case**, switch to the **Lifecycle** tab see your prompt template lifecycle being tracked. If it doesn't appear immediately refresh your screen.

The screenshot shows the AI use cases interface for an insurance claim summarization project. The Lifecycle tab is selected. On the left, there's a sidebar with 'Welcome to your new AI use case' and 'Discover next steps'. The main area has two sections: 'Organize team work with approaches' (with icons for circles, triangles, and squares) and 'Manage AI assets with versions' (with icons for arrows). Below these are sections for 'Development', 'Validation', and 'Operation'. Under 'Development', the 'Default approach' is selected, showing 'Latest: 0.0.1 | 1 version'. The validation section says 'No models promoted to a pre-production space.' and the operation section says 'No models promoted to a production space.' There are also 'Show deleted assets' and 'New approach +' buttons.

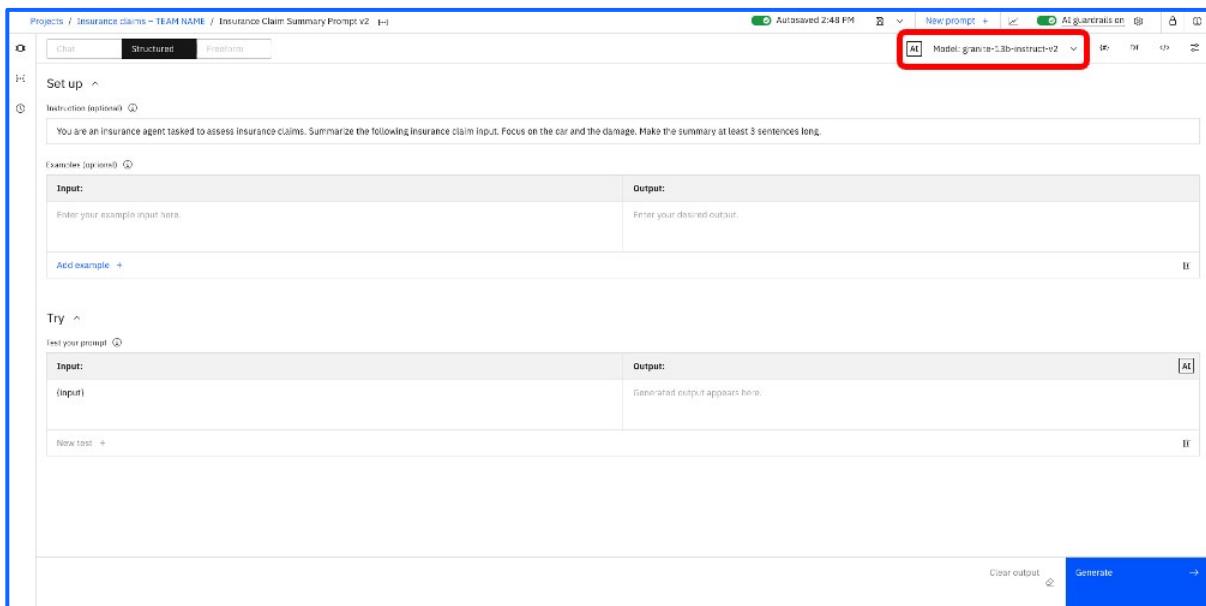
This is where you can view any changes to the prompt template as we continue with the lab.

Update your prompt template

We will now make additional changes to the prompt template, which we'll do by creating a new version of the original prompt.

For that, use the main menu to navigate to your project. We'll create a new prompt that is a copy of the first one. Do that by opening your template from the **Assets** tab and use **Save as** to give it a new name, such as “Insurance Claim Summary Prompt v2”. That opens it for editing again.

Change the model to “**granite-13b-instruct-v2**” to introduce a change in this new version of the prompt template!



Evaluate your prompt template by following the same steps as before ([Evaluating your prompt](#)).

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

[Track in AI use case](#)

[Learn more](#)

Once you have completed the evaluation of this new prompt template, you can track it in your AI use case. Go to your project by selecting the project name in the path at the top.

Track it in your AI use case by selecting **Minor Change (version 0.1.0)**.

Approach: Default approach | Use case: Insurance claim summarisation - TEAM NAME
You are updating version 0.0.1. Choose the type of change that describes the update to adjust the version number accordingly.

assign version

Minor change

Your model's features have been extended but its structure has remained unchanged so that interfaces don't need to be adapted before it can be used again.

0.0.1 → 0.1.0

Patch change

Your model has received a small fix or a dependency update which doesn't affect its structure or features.

0.0.1 → 0.0.2

Major change

Your model's structure has changed and interfaces have to be adapted before it can be used again.

0.0.1 → 1.0.0

Custom

Assign a custom version that reflects the scope of changes to the model.

Version number
0.1.0

Comment (optional)

Provide context for this change. Comments are saved in the model submit to govern.

[Cancel](#) [Back](#) [Next](#)

Check that it appears in the **Lifecycle** tab of your AI use case.

Welcome to your new AI use case

Default approach

A default approach for tracking your AI assets.

Development

Experiment with different approaches of ML development or prompt engineering and test them in projects.

v0.1.0

- Insurance claims – TEAM NAME
 - Insurance Claim Summary Prompt v2
- No models promoted to a pre-production space.
- No models promoted to a production space.

v0.0.1

- Insurance claims – TEAM NAME
 - Insurance Claim Summary Prompt
- No models promoted to a pre-production space.
- No models promoted to a production space.

You will see that the two prompts are now tracked separately within the same AI use case, making it easy to track different prompts, models or parameters for a single use case/application.

Evaluation Studio

The purpose of the Evaluation Studio is to look at different prompts and identify which is the strongest performing. For the lab, you will compare the two prompt templates you've created. Note if you didn't get this far, then there are existing prompts you can use. The following lab explains how simple it is to perform an evaluation using the Evaluation Studio.

Navigate back to your project and select “New Asset”.

Name	Last modified	Actions
Insurance claim summarization Prompt template	4 days ago Modified by Service	⋮
Insurance Claims Summarisation Attempt 3 Prompt template	1 week ago Modified by Service	⋮
Insurance Claims Summarisation Attempt 1 Prompt template	1 week ago Modified by Service	⋮
Insurance Claims Summarisation Attempt 2 Prompt template	2 weeks ago Modified by Service	⋮
Insurance claim key information extraction Prompt template	2 weeks ago Modified by Service	⋮
Insurance claim suggested next steps Prompt template	2 weeks ago Modified by Service	⋮

In the new UI, search for Evaluate and Compare prompts.

The screenshot shows the IBM Watsonx interface with the title bar "IBM Watsonx" and "Projects / Team 1". A sidebar on the left lists "Prepare data", "Work with models", and "Automate model lifecycles". The main area has a heading "What do you want to do?" with the sub-instruction "Select a task based on your goal. You'll use a tool to create an asset for that goal." Below this is a search bar "Search for a task or tool". Under "Recent tasks", there are two items: "Evaluate and compare prompts" (with Evaluation Studio) and "Work with data and models in Python or R notebooks" (with Jupyter notebook editor). A section titled "Prepare data" contains four options: "Connect to a data source" (with Connection), "Ground gen AI with vectorized documents" (with Vector indexes), "Prepare and visualize data" (with Data Refinery), and "Define reusable sets of parameters" (with Parameter set).

Leave the name and description as it. Select the task type as Summarisation

The screenshot shows the "Evaluate and compare prompts" setup dialog. On the left, a sidebar lists "Setup", "Prompt templates", "Metrics", "Test data", and "Review and run". The main area has a "Setup" section with a sub-instruction "Define settings for evaluating and comparing prompts. You can only compare prompt templates of the same task type." and a "Learn more" link. It includes fields for "Name" (set to "Evaluation experiment") and "Description (optional)" (set to "Enter a description"). Below these is a "Select the task type" section with a grid of icons: Q&A, Summarization (which is selected and highlighted in blue), RAG, Classification, Generation, and Extraction. At the bottom are "Cancel", "Back", "Next" (which is highlighted in blue), and "Cookie Preferences".

Select the two different prompts you created earlier. In this example they are labelled Insurance Claims Summarisation Attempt 1 and 2. Click Next.

Evaluate and compare prompts

Prompt templates

Select at least 2 prompt templates of the specified task type from your project. Prompt templates you compare must have the same variables.

Prompt template	Foundation model	Variables
<input checked="" type="checkbox"/> Insurance Claims Summarisation Attempt 1	google/flan-ul2	<input type="button" value="input"/>
<input checked="" type="checkbox"/> Insurance Claims Summarisation Attempt 2	google/flan-t5-xxl	<input type="button" value="input"/>
<input type="checkbox"/> Insurance Claims Summarisation Attempt 3	ibm/granite-3-Bb-instruct	<input type="button" value="input"/>
<input type="checkbox"/> Insurance claim summarization	google/flan-ul2	<input type="button" value="input"/>

Cancel Back Next Cookie Preferences

Review the different metrics which you are evaluating against. Again these fall under the categories General AI Quality and Model Health and can be configured. .

Evaluate and compare prompts

Metrics

Select metric groups for evaluating the selected prompt templates. You see metrics that are available for the specified task type.

[Learn more](#)

Generative AI Quality

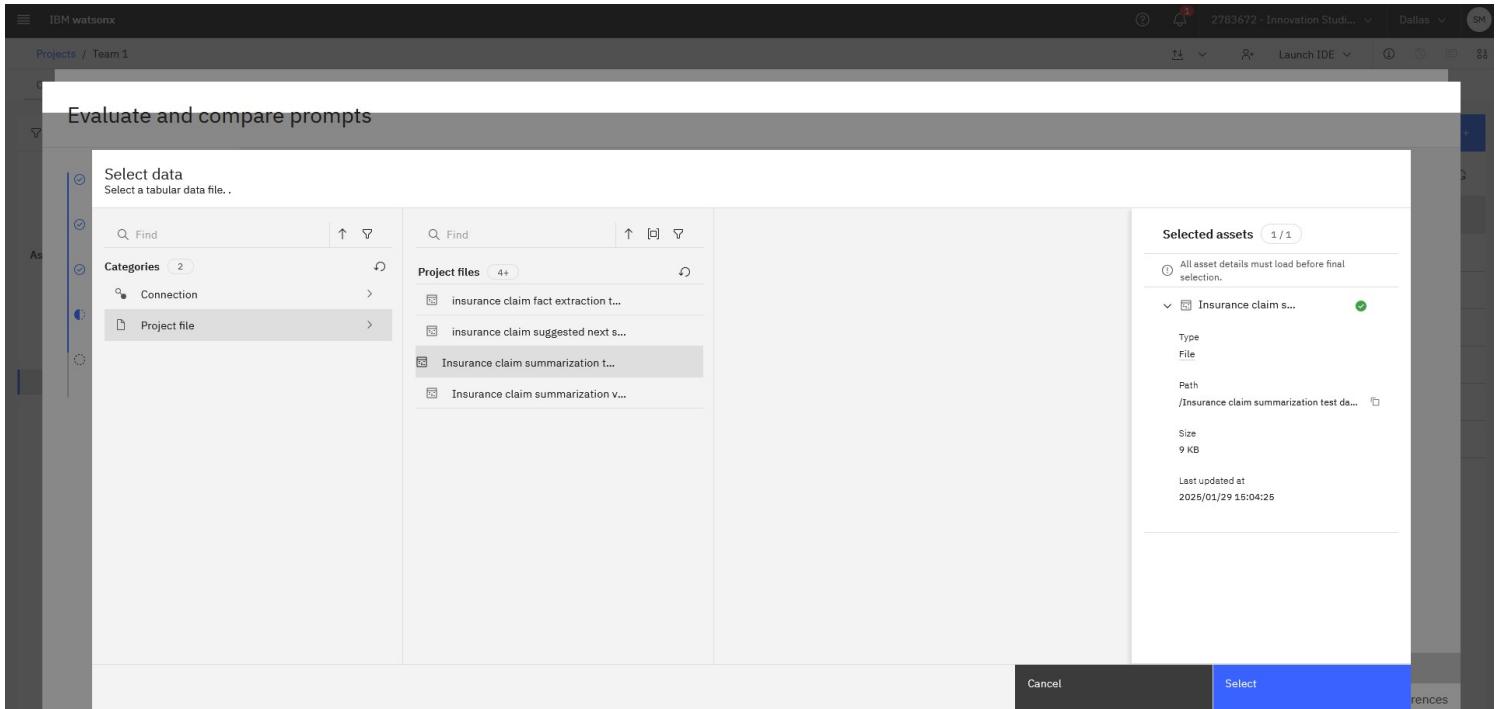
The Generative AI Quality monitor calculates a variety of metrics based on prompt template task type. Some metrics compare model output to the reference output you provide. Other metrics analyze model input and output and do not require reference output.

[Configure](#)

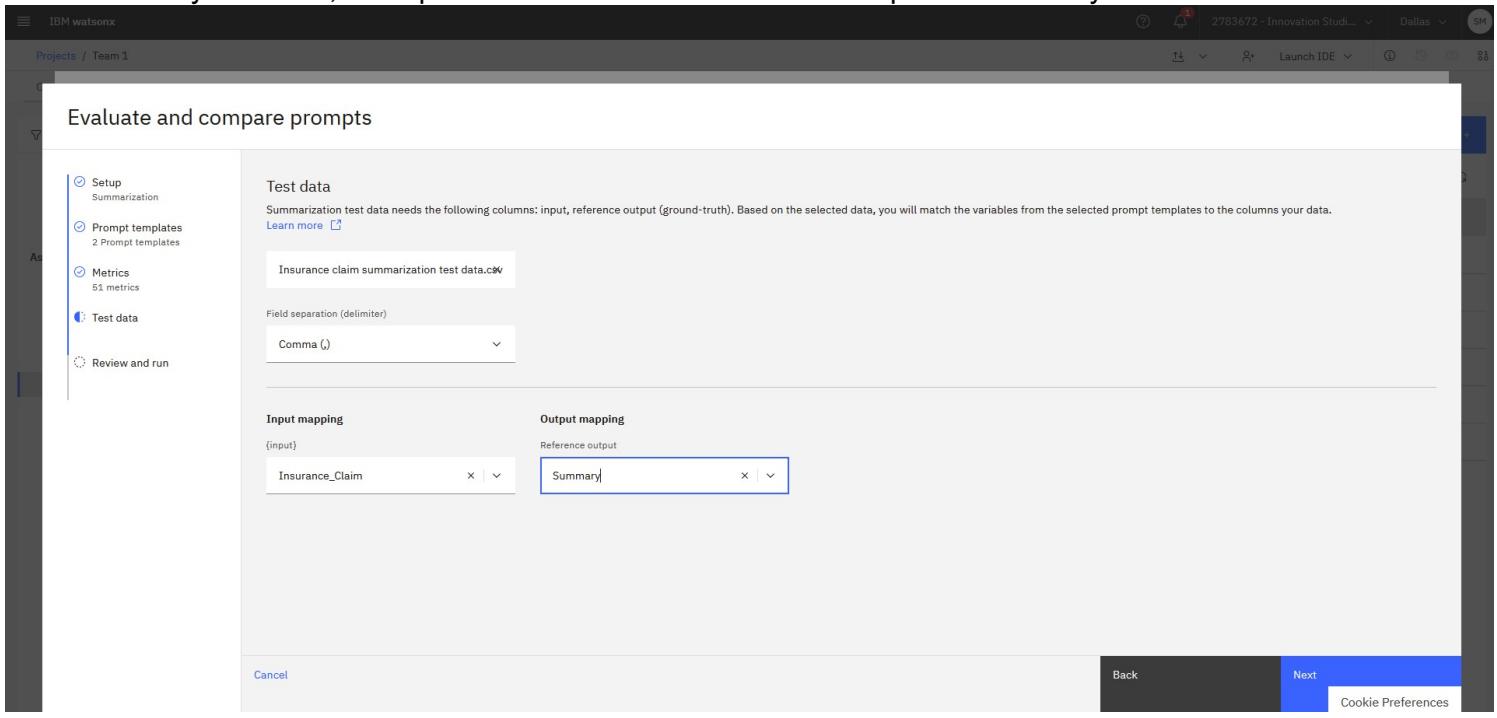
Group	Included metrics
Content Analysis	Coverage, Density, Compression, Repetitiveness, Abstractness
ROUGE	ROUGE-1, ROUGE-2, ROUGE-L, ROUGE-Lsum
SARI	SARI
METEOR	METEOR
Text quality	F1 Score, Precision, Recall
BLEU	BLEU

Cancel Back Next Cookie Preferences

On the Test Data screen, please select the Insurance Claims Summarisation test data csv from the project. This is under the project file



Similarly to before, the Input is “Insurance Claim” and the Output is “Summary”



Review the evaluation before running it, this will take a few minutes.

IBM Watsonx

Projects / Team 1 / Evaluation experiment

Evaluation in progress...

You can continue to work on other things while your evaluation is in progress.

[View jobs](#)

Test data as input
Each record of your test data serves as input for the prompt template you are evaluating.

Evaluation of output
Generated output is evaluated with specified metrics. Some metrics compare the generated output with reference data (ground truth).

Result comparison
After the evaluation completes, you can compare results to decide which prompt template is best for your use case.

Spend some time going through the results, identifying how each attempt performs differently across several criteria.

IBM Watsonx

Projects / Team 1 / Evaluation experiment

Metrics comparison

Evaluation details

- Task type: Text summarization
- Test data: Insurance claim summarization t...
- Metrics: Model health (33 metrics), Generative AI quality (18 metrics)

Prompt template	Compression	Repetitiveness	Coverage	Density	Abstractness
Insurance Claims Summarisatik google/ftan-t5-xxd	0.009	0.073	0.37	0.211	0.197
Insurance Claims Summarisatik ibm/gptne-3-0b-instruct	1.732	0.002	0.242	0.004	0.062
Insurance claim summarization google/ftan-uf2	5.283	0.085	0.477	0.406	0.043

Content Analysis

Once done, navigate back to the project nad you will see these results exist as an evaluation result.

Phase 4: Deployment

If we consider that you've iterated the prompts and found the right combination of model, parameters, model variables until you're happy. You'd now look to promote that asset into production. When you put an asset in production, there are a different set of Risks associated with it for example Drift. This next part of lab will show you how to promote the asset into production and the features available plus turning on the evaluation for drift.

Deployment Space

A prompt template is only useful if it is accessible by an application; this prompt template needs to be made available in a consumable way. What's more, the deployment process needs to be completed in an orderly manner. For example, you cannot allow everyone to put prompt templates/models into production. For this purpose, the watsonx.ai environment includes deployment spaces.

Anyone can create their own personal deployment space and promote prompt templates/ models to it. However, in an enterprise, designated administrators should be the only ones able to promote a model to a production deployment space.

Go back to your Use Case and scroll down to Workspaces, click to add a Workspace in Operate:

Overview Lifecycle Access

Additional Details ⓘ
Data not available from OpenPages

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

Development
Team 5

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 +](#)



The diagram illustrates the deployment hierarchy. At the center is the "AI use case". Surrounding it are three concentric circles representing deployment spaces: "Workspaces for Develop" (top), "Workspaces for Operate" (bottom-left), and "Workspaces for Validate" (bottom-right). A blue arrow points from the "AI use case" towards the "Workspaces for Develop" circle. Dashed lines connect the "AI use case" to the other two circles, indicating they are also associated.

Click on the button shown below to create a [New space](#).

Space
A space can be associated to multiple AI use cases.

New space +

Name	Created	Your role	Tracked AI assets	Stage	Associates
------	---------	-----------	-------------------	-------	------------

You will see the window shown on the next page.

Fill in the details as specified below:

- Name the new deployment space something appropriate
- Select the storage service and machine learning service from the options that are available under the respective dropdown menus on the right side
- Select the Deployment stage as 'Production'
- Press **Create**.

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

Define details	Select services
Name <input type="text" value="Deployment Space - Production"/>	Select storage service <small>①</small> <input type="text" value="Cloud Object Storage-pj"/>
Description (Optional) <input type="text" value="Deployment space description"/>	Select watsonx.ai Runtime service (optional) <small>①</small> <input type="text" value="Watson Machine Learning-a1"/>
Deployment stage <small>①</small> <input type="text" value="Production"/>	Upload space assets (optional) <small>Populate your space with assets exported from a project or space to a .zip file. You can add more assets after the space is created.</small> <input type="text" value="Drop .zip file here or browse your files to upload"/>
Deployment space tags (optional) <small>①</small> <input type="text" value="Find or create tags"/>	
Advanced Settings <input type="text" value=""/>	

Cancel Create

Once you have created this deployment space, press **Save** to associate it to the workspace **Operate**.



Next, find your second prompt version (the one created during the Develop phase) within **Projects** and promote it to this new workspace. Select the Deployment space you created earlier.

We then need to deploy the prompt, click on the 3 buttons next to your asset and choose deploy:

To confirm the steps worked, you should see the prompt template in the *Production deployment* space on the far-right panel, with the *Pending Evaluation* label

A default approach for tracking your AI assets.

Development	Validation	Operation
Experiment with different approaches of ML development or prompt engineering and test them in projects.	Evaluate models and prompt templates with validation data.	Evaluate and monitor models and prompt templates in production.

v0.1.0

- Insurance claims – TEAM NAME
 - Insurance Claim Summary Prompt v2
- Deployment Space - Validation
 - Insurance Claim Summary Prompt v2
 - Insurance Summarisation - Deployment
- Deployment Space - Production
 - Insurance Claim Summary Prompt v2
 - Insurance Summarisation - Prod Deployment

Well done, you have successfully deployed your first asset into production.

Managing Assets in production

Click on the recently created production deployment space (the one that is shown to be pending evaluation), and click [Open in Space](#).

Deployment Space - Production

c60cffa0-e16f-44ce-95a6-2aa8f6a7a7bf

Created by Tasos Christodoulou	Created 29 January 2025 at 09:56:03	Asset name Insurance Claim Summary Prompt v2
Number of deployments 1		

(i) Deployment Space - Production > Insurance Summarisation - Prod Deployment

Deployed ✓ | Online | 76a227b1-4dfb-4754-a661-2e358847a307

Scoring url
<https://us-south.ml.cloud.ibm.com/ml/v1/deployments/76a227b1-4dfb-4754-a661-2e358847a307/text/generation>
https://us-south.ml.cloud.ibm.com/ml/v1/deployments/76a227b1-4dfb-4754-a661-2e358847a307/text/generation_stream

Evaluation activity

Threshold alerts
0

01

Cancel **Open in space**

The first thing to talk through is the API reference. As discussed, there is no point in developing a model if it isn't easy to access. With the deployment space, you can see that there are public and private API endpoints to allow an application to call the model. Making it even simpler, there are code snippets available in multiple languages to allow this to be baked into the application code. Thus, making our production asset readily accessible.

The screenshot shows the IBM WatsonX interface for managing deployments. The current view is for the 'sam prod deployment'. The 'Test' tab is active, displaying information about endpoints and code snippets for testing the deployment.

Endpoints for inferencing:

- Private endpoint:**
 - Text endpoint URL: https://private.us-south.ml.cloud.ibm.com/v1/deployments/sam_prod_deploy
 - Stream endpoint URL: https://private.us-south.ml.cloud.ibm.com/v1/deployments/sam_prod_deploy
 - Bearer <token>:
- Public endpoint:**
 - Text endpoint URL: https://us-south.ml.cloud.ibm.com/v1/deployments/sam_prod_deploy
 - Stream endpoint URL: https://us-south.ml.cloud.ibm.com/v1/deployments/sam_prod_deploy

Code snippets:

```
# NOTE: you must set $API_KEY below using information retrieved from your IBM Cloud account (https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/ml-authentication.html?context=ex)
curl --insecure -X POST --header "Content-Type: application/x-www-form-urlencoded" --header "Accept: \napplication/json" --data-urlencode "grant_type=urn:ibm:params:oauth:grant-type:apikey" \
--data-urlencode "apikey=$API_KEY" "https://iam.cloud.ibm.com/identity/token"
```

the above CURL request will return an auth token that you will use as \$IAM_TOKEN in the scoring request below
To score the model with the curl command below, run the following command
curl -X POST -H "Content-Type: application/json" -H "Accept: application/json" -H "Authorization: \nBearer \$IAM_TOKEN" -d "[{"parameters": [{"input": ""}]}]" "https://private.us-south.ml.cloud.ibm.com/v1/deployments/sam_prod_deploy/text/generation_stream?version=2021-08-01"

About this deployment:

- Name:** sam prod deployment
- Description:** No description provided.
- Deployment Details:**
 - Deployment ID: 98cc28ed-8164-4b...
 - Serving name: sam_prod_deploy
 - Copies: 1
- Tags:** Add tags to make assets easier to find.
- Associated asset:** [sam 2](#) e745fact-e7fd-48b7-b8fb-a66b62ee8688
- Last modified:** 26 minutes ago
- Created on:** Feb 11, 2025

Now, we can test it to make sure it is functioning as expected.

Navigate to the **Test** tab within this **Deployment Space** to ensure the function is working as expected.

The screenshot shows the IBM WatsonX interface for managing deployments. The current view is for the 'Insurance Summarisation - Prod Deployment'. The 'Test' tab is active, displaying deployment details, test results, and various performance metrics.

Deployment details:

- Deployment ID: 98cc28ed-8164-4b...
- Serving name: Insurance Summarisation - Prod Deployment
- Deployment status: Deployed

Test details:

- Tests run: 0
- Tests passed: 0
- Tests failed: 0

Model health:

- Records: 0
- Total input token count: 0
- Total output token count: 0
- Latency (record): 0 ms (Median record latency)

Within this tab, click on the **Generate** button to make sure the prompt works.

The screenshot shows the 'Insurance Summarisation - Prod Deployment' page. At the top, there are tabs for 'API reference', 'Test' (which is selected), 'Evaluations', and 'AI Factsheet'. Below the tabs, a section titled 'Enter input data' has three tabs: 'Text' (selected), 'Stream', and 'JSON'. A note says: 'These variables were specified in the Prompt Lab to provide values for the prompt template. You can edit the value of these variables before generating text.' Under 'Prompt', there is a text box containing: 'You are an insurance agent tasked to assess insurance claims. Summarize the following insurance claim input. Focus on the car and the damage. Make the summary at least 3 sentences long.' Below this are 'Input:' and 'Output:' fields. In the bottom right corner of the main content area, there is a blue 'Generate' button with a red box around it.

Review the schemas of the payload and the feedback data. The datasets (csv) files need to follow these schemas when we upload them.

The screenshot shows the 'Prompt results' section. It says: 'Choose a display format for viewing the prompt results.' with 'Table view' selected. A table displays the following data:

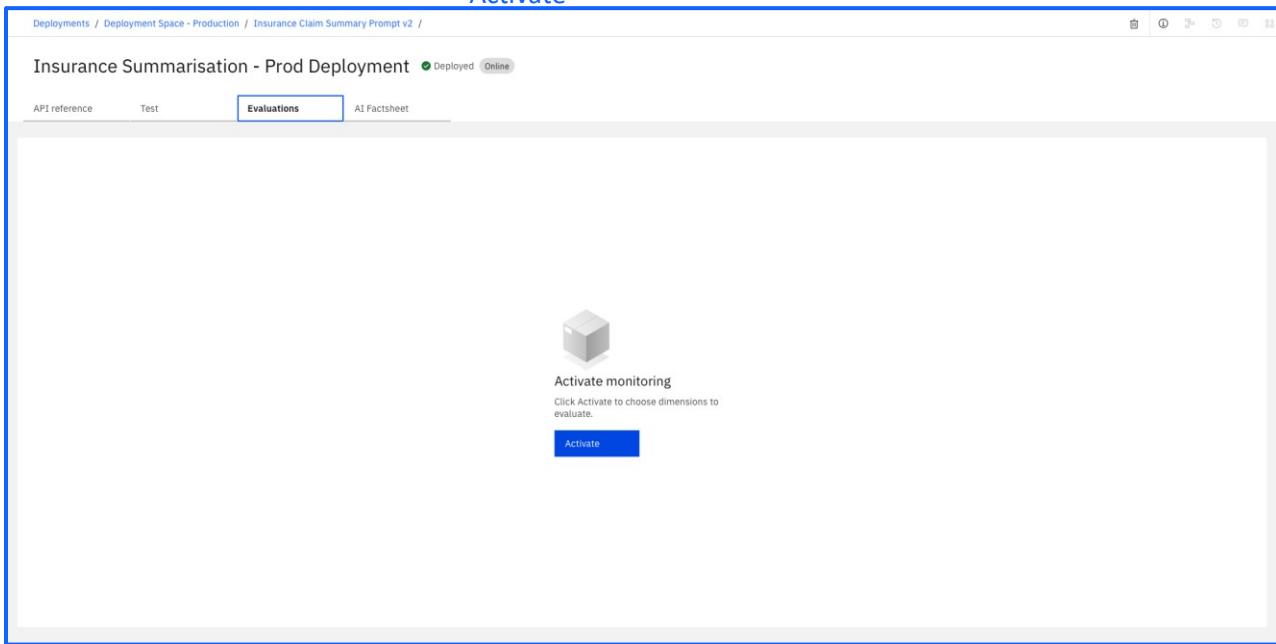
Key	Value
1 Generated text	Car A hit Car B from behind, totaling Car B.
2 Generated token count	14
3 Input token count	45
4 Stop reason	End of sequence token encountered
5 Seed	0

In the bottom right corner of the table view, there is a blue 'Download JSON file' button with a red box around it.

We are now ready to prime the prompt deployment with baseline data.

Key	Value
<i>Generated text</i>	Car A hit Car B from behind, totalling Car B.
<i>Generated token count</i>	14
<i>Input token count</i>	45
<i>Stop reason</i>	End of sequence token encountered
<i>Seed</i>	0

Go to the **Evaluations** tab and click the [Activate](#) button.



You should notice that now you have activated the monitoring, you will automatically have a new dimension called **Drift v2** as an additional option. As shown in the screenshot below, the description gives details as to how it measures distribution shifts between training and payload data – which we must now upload.

Evaluate prompt template

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

Select dimensions	Select dimensions to evaluate														
<input checked="" type="checkbox"/> Review and evaluate	These dimensions are based on the prompt template task type. Learn more														
	Column names in payload data must match the prompt variable names in the prompt template. Feedback data must include an additional label column for reference output. Provide the name of the label column in the field below. Note that the Drift v2 evaluation requires 100+ records to establish a baseline.														
	Label column <input type="text" value="reference_output"/>														
	Advanced settings														
	<table border="1"> <thead> <tr> <th>Dimension</th> <th>Description</th> </tr> </thead> <tbody> <tr> <td><input checked="" type="checkbox"/> Drift v2</td> <td>Computes distribution shifts between training and payload data.</td> </tr> <tr> <td><input type="checkbox"/> Embedding drift</td> <td>Embedding drift is defined as the percentage of records where the runtime data was detected as an outlier as compared to the baseline data. This is an opt-in experience for the user. Once enabled, the user needs to provide embeddings along with the data for this metric to be computed.</td> </tr> <tr> <td><input type="checkbox"/> Output drift</td> <td>Output drift is defined as the change in distribution for the model output between the baseline data and the runtime data. For classification problems, this is computed on the individual class probabilities. For regression problems, this is computed on the model prediction. For the generative AI models, this is computed on the log probability of the final generated text.</td> </tr> <tr> <td><input type="checkbox"/> Input metadata drift</td> <td>Input metadata drift is defined as the change in distribution for the model input metadata between the baseline data and the runtime data. If the tag, 'field_type', is present and set to 'meta', it will denote drift in the input metadata column in 'field_name' tag. Otherwise, this metric is at subscription level.</td> </tr> <tr> <td><input type="checkbox"/> Output metadata drift</td> <td>Output metadata drift is defined as the change in distribution for the model output metadata between the baseline data and the runtime data. If the tag, 'field_type', is present and set to 'meta', it will denote drift in the output metadata column in 'field_name' tag. Otherwise, this metric is at subscription level.</td> </tr> <tr> <td><input type="checkbox"/> Generative AI Quality</td> <td>The Generative AI Quality monitor calculates a variety of metrics based on prompt template task type. Some metrics compare model output to the reference output you provide. Other metrics analyze model input and output and do not require reference output.</td> </tr> </tbody> </table>	Dimension	Description	<input checked="" type="checkbox"/> Drift v2	Computes distribution shifts between training and payload data.	<input type="checkbox"/> Embedding drift	Embedding drift is defined as the percentage of records where the runtime data was detected as an outlier as compared to the baseline data. This is an opt-in experience for the user. Once enabled, the user needs to provide embeddings along with the data for this metric to be computed.	<input type="checkbox"/> Output drift	Output drift is defined as the change in distribution for the model output between the baseline data and the runtime data. For classification problems, this is computed on the individual class probabilities. For regression problems, this is computed on the model prediction. For the generative AI models, this is computed on the log probability of the final generated text.	<input type="checkbox"/> Input metadata drift	Input metadata drift is defined as the change in distribution for the model input metadata between the baseline data and the runtime data. If the tag, 'field_type', is present and set to 'meta', it will denote drift in the input metadata column in 'field_name' tag. Otherwise, this metric is at subscription level.	<input type="checkbox"/> Output metadata drift	Output metadata drift is defined as the change in distribution for the model output metadata between the baseline data and the runtime data. If the tag, 'field_type', is present and set to 'meta', it will denote drift in the output metadata column in 'field_name' tag. Otherwise, this metric is at subscription level.	<input type="checkbox"/> Generative AI Quality	The Generative AI Quality monitor calculates a variety of metrics based on prompt template task type. Some metrics compare model output to the reference output you provide. Other metrics analyze model input and output and do not require reference output.
Dimension	Description														
<input checked="" type="checkbox"/> Drift v2	Computes distribution shifts between training and payload data.														
<input type="checkbox"/> Embedding drift	Embedding drift is defined as the percentage of records where the runtime data was detected as an outlier as compared to the baseline data. This is an opt-in experience for the user. Once enabled, the user needs to provide embeddings along with the data for this metric to be computed.														
<input type="checkbox"/> Output drift	Output drift is defined as the change in distribution for the model output between the baseline data and the runtime data. For classification problems, this is computed on the individual class probabilities. For regression problems, this is computed on the model prediction. For the generative AI models, this is computed on the log probability of the final generated text.														
<input type="checkbox"/> Input metadata drift	Input metadata drift is defined as the change in distribution for the model input metadata between the baseline data and the runtime data. If the tag, 'field_type', is present and set to 'meta', it will denote drift in the input metadata column in 'field_name' tag. Otherwise, this metric is at subscription level.														
<input type="checkbox"/> Output metadata drift	Output metadata drift is defined as the change in distribution for the model output metadata between the baseline data and the runtime data. If the tag, 'field_type', is present and set to 'meta', it will denote drift in the output metadata column in 'field_name' tag. Otherwise, this metric is at subscription level.														
<input type="checkbox"/> Generative AI Quality	The Generative AI Quality monitor calculates a variety of metrics based on prompt template task type. Some metrics compare model output to the reference output you provide. Other metrics analyze model input and output and do not require reference output.														

[Cancel](#) [Back](#) [Next](#)

Once you have walked through the steps, click **Activate** to activate monitoring of the prompt in production. This asset is now being monitored in production on a scheduled basis, metrics captured will be sent to the Governance Console and tracked within the Use Case you created earlier.

Note: The Prompt Results that you see in the lab may differ slightly to what you see above.

IBM WatsonX

Deployments / Deployment Space - Production / Insurance Claim Summary Prompt v2 /

Insurance Summarisation - Prod Deployment Deployed Online

Evaluations API reference Test AI Factsheet

Last evaluation: --

Deployment details	Test details	Model health	Actions
	<p>0 Tests run</p> <p>Tests passed: 0 Tests failed: 0</p>	<p>Records: 0</p> <p>Latency (record): 0 ms</p> <p>Total input token count: 0</p> <p>Total output token count: 0</p>	Evaluate now Compare Configure monitors View model information

Generative AI Quality - Text summarization

! Not evaluated

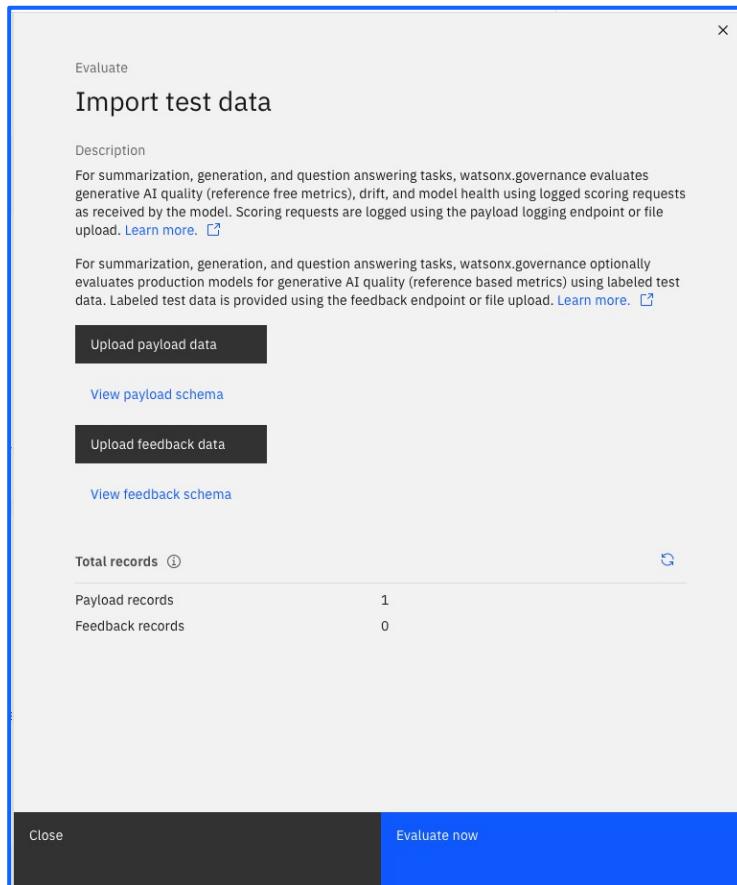
Awaiting model transactions to evaluate.
Evaluation will run when the minimum number of transactions is met.

Drift Monitoring (Optional – feel free to progress to the [Governance Console](#) section)

To demonstrate drift we need to establish a baseline. Consider this as the ground truth for the data. That is to say at time of production the data is up to date and accurate. Over time of course, Drift will be introduced as the data becomes outdated and no longer relevant. The calculation of this works off the baseline data.

Select Upload Payload data and then upload the file called:

Insurance claim summarisation drift payload 0 - baseline.csv



Note: We are not uploading any feedback data at this point.

The screenshot shows the IBM WatsonX interface for a deployment named "Insurance Summarisation - Prod Deployment". The "Evaluations" tab is active. A modal window titled "Import test data" is displayed, showing a success message: "Insurance claim summarization drift payload 0 - baseline.csv uploaded successfully OK". The main dashboard displays deployment details, test results (0 tests run, 0 passed, 1 failed), and model health metrics (records: 101, token count: 342, total input token count: 55). A note states "Not evaluated" due to awaiting model transactions.

Once that has been successfully uploaded, Click on the [Evaluate now](#) button.

The **Evaluation** will fail due to a violation of the **Coverage Metric**. We could of course amend the threshold of the metric for our **Evaluation** to Pass, however for the purposes of our workshop will accept it as it is.

The screenshot shows the IBM WatsonX interface for the same deployment. The "Evaluations" tab is active. The test details now show 1 test run, 0 passed, and 1 failed. The model health metrics remain the same: 101 records, 342 total input token count, and 55 total output token count. The latency is listed as 1.634 ms.

If you now click on the [Configure monitors](#) button in the [Actions](#) dropdown menu, you will see **Drift v2** enabled.

As shown in the screenshot below, you will have a baseline of 100 records. You can edit the threshold values here as well which would be important in a real-world scenario, to adjust them to an appropriate level given the use case and prompt.

For now, we'll keep the default values.

Insurance Summarisation - Prod Deployment

Model info

Model details

Evaluations

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

Drift v2

Description

The Drift monitor checks if your deployments are up-to-date and behaving consistently. Model input/output data is analyzed in relation to the training/baseline data.

Compute the drift archive

Compute option

Compute in Watson OpenScale

Baseline data

Number of records

100

Sample size

Minimum sample size

10

Maximum sample size

1,000

Drift thresholds

Upper thresholds

Output drift

0.05

Input metadata drift

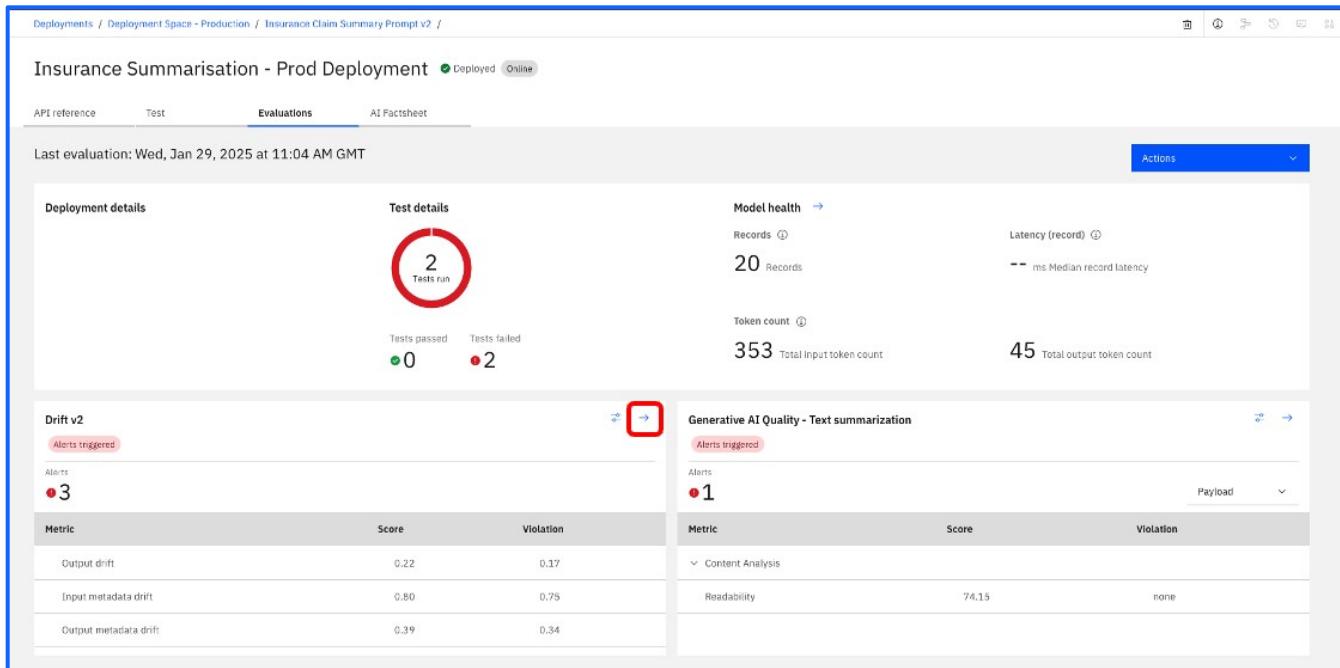
0.05

The system will now monitor invocations of the prompt and detect any potential drift. We will simulate an example of this by uploading the following datasets.

- Insurance claim summarisation drift payload 1.csv
- Insurance claim summarisation drift payload 2.csv
- Insurance claim summarisation drift payload 3.csv
- Insurance claim summarisation drift payload 4.csv

Upload all of the csv files as payload data, but be sure to let some time pass between each run, to be sure that the evaluation shows a meaningful timeline (and drift).

Once you have run these evaluations, you will see a sub-tab under **Evaluations** to show a diagram for the drift of input and output. Click on the  blue arrow button in the corner of the Drift v2 section to get a more detailed view.



The screenshot shows the IBM Data Platform interface for a deployment named "Insurance Summarisation - Prod Deployment". The top navigation bar includes "Deployments", "Deployment Space - Production", "Insurance Claim Summary Prompt v2", and "Actions". Below the navigation is a toolbar with "API reference", "Test", "Evaluations" (which is selected), and "AI Factsheet". A status indicator shows "Deployed" and "Online".

The main content area displays deployment details: "Last evaluation: Wed, Jan 29, 2025 at 11:04 AM GMT". It also shows test details with a circular progress bar containing the number "2" and the text "Tests run". Below this, "Deployment details" show "Tests passed: 0" and "Tests failed: 2".

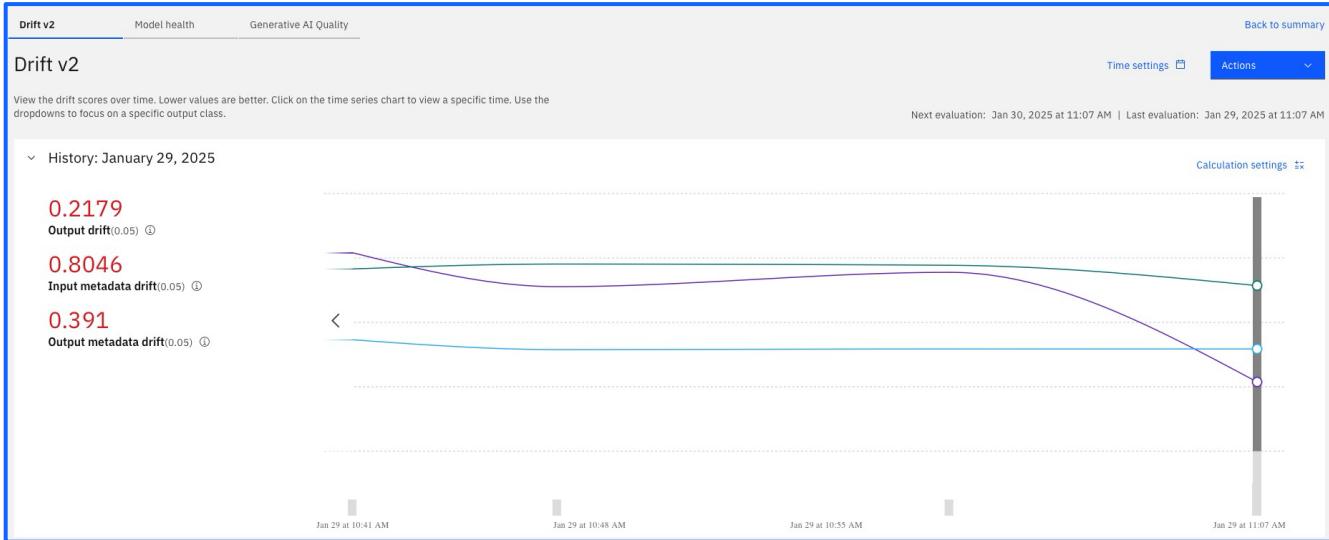
Drift v2 section (highlighted with a red box around the "→" button):

Metric	Score	Violation
Output drift	0.22	0.17
Input metadata drift	0.80	0.75
Output metadata drift	0.39	0.34

Generative AI Quality - Text summarization section:

Metric	Score	Payload
Content Analysis	74.15	none

Make sure you have selected the right time settings to see the drift change over time. If you have any issues, ask for assistance from your facilitators.



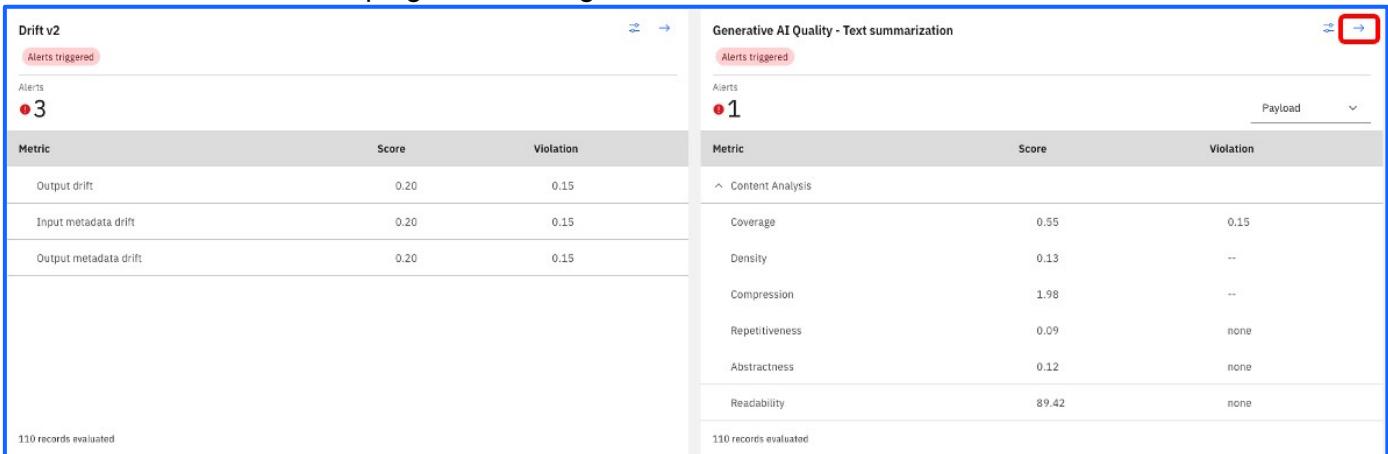
There are only three criteria which go into the drift calculations: *output*, *output metadata drift* and *input metadata drift*.

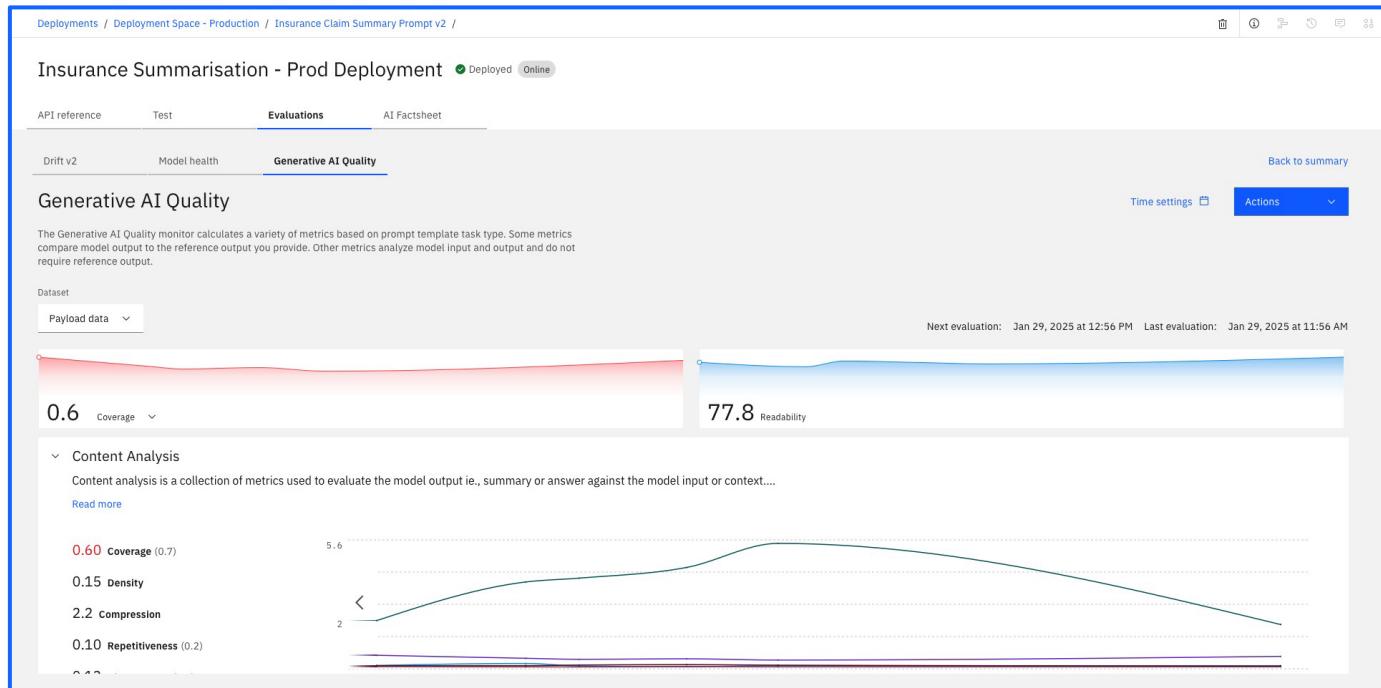
These metrics do not offer any information regarding the *quality* of the prompt – In order to measure that, we need to upload *feedback* data.

Navigate back to the **Evaluations** tab and once again click on the [Evaluate now](#) button within the [Actions](#) dropdown menu. Upload the following file:

- Insurance claim summarisation feedback.csv

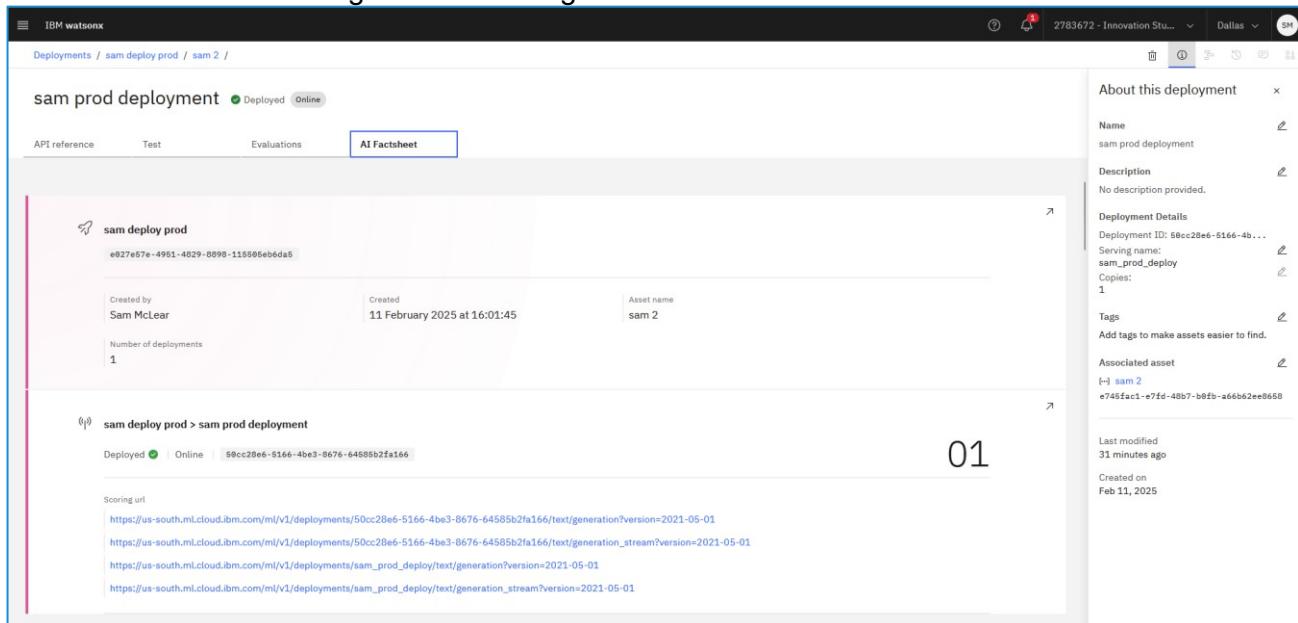
When this run is finished, you will see a panel with **Generative AI Quality - Text summarisation** metrics on the lower right of the screen, next to the Drift v2 section. If the option allows, click on the blue arrow button in the top right corner to get a more detailed view of the metrics collected.





Finally, we are ready to advance the status of the use case, to identify it as a model in **Operation**.

The final tab, AI Factsheet, captures all of the facts about the AI Asset. It is also possible to view information such as the serving name on the right hand side of the screen.



The prompt is deployed in production and is being monitored for both drift and quality. As seen back in the lifecycle screen.

The screenshot shows the IBM Watsonx interface for managing AI use cases. The top navigation bar includes 'AI use cases / Insurance claim summarisation – TEAM NAME', 'Overview', 'Lifecycle' (which is selected), 'Access', and 'Export report'. The main content area has three tabs: 'Welcome to your new AI use case', 'Discover next steps', and 'Organize team work with approaches'. The 'Organize team work with approaches' tab is active, displaying a diagram of four interconnected circles labeled 'Approach A', 'Approach B', 'Approach C', and 'Approach D'. Below the diagram, a section titled 'Manage AI assets with versions' explains that each approach starts a new version series for identifying associated assets. A 'Default approach' is listed with version v0.1.0 (2 versions). The interface is divided into 'Development', 'Validation', and 'Operation' sections. Under Development, there's a note about experimenting with different approaches. Under Validation, it says 'Evaluate models and prompt templates with validation data.' Under Operation, it says 'Evaluate and monitor models and prompt templates in production.' At the bottom, there are sections for 'v0.1.0' and 'v0.0.1', each with a summary of deployment spaces and their status.

Welcome to your new AI use case

Discover next steps

Organize team work with approaches

Manage AI assets with versions

Approach A

Approach B

Approach C

Approach D

Default approach

Latest: 0.1.0 | 2 versions

A default approach for tracking your AI assets.

Development

Experiment with different approaches of ML development or prompt engineering and test them in projects.

Validation

Evaluate models and prompt templates with validation data.

Operation

Evaluate and monitor models and prompt templates in production.

v0.1.0

Insurance claims – TEAM NAME

Insurance Claim Summary Prompt v2. 1

Deployment Space - Validation

Insurance Claim Summary Prompt v2

Insurance Summarisation - Deployment. 1

Evaluated

Deployment Space - Production

Insurance Claim Summary Prompt v2

Insurance Summarisation - Prod Deployment. 1

Evaluated

v0.0.1

Insurance claims – TEAM NAME

No models promoted to a pre-production space.

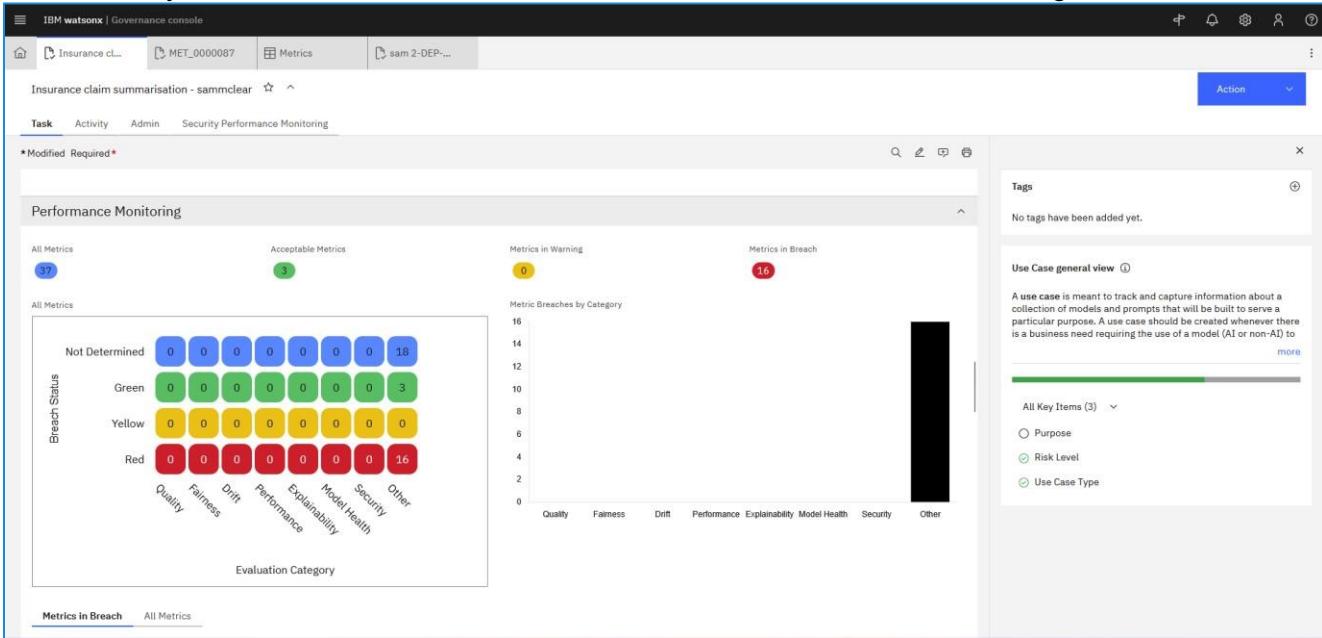
No models promoted to a production space.

Governance Console

For the last part of the lab, you will see how the work carried out throughout this lab was automatically captured in the Use Case on the Governance Console.

Navigate back to the Governance Console either from the Use Case view or [this](#) link.

Click back into your teams' Use Case and scroll down to the Performance Monitoring section:



Scroll down to the Relationships screen. You can see the different objects that are now tied to the Use Case. To drill down further into the models themselves and their performance, click on the name and additional objects will appear.

IBM Watsonx | Governance console

Insurance claim summarisation - sammclear

Task Activity Admin Security Performance Monitoring

* Modified Required*

No results

Relationships

Primary Parent Parent Child

Administration

Tags

No tags have been added yet.

Use Case general view

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

All Key Items (3)

Purpose Risk Level Use Case Type

Related Models

All Models Prompts and Tuners Associated Foundation Models Deployments Model Links Model Groups

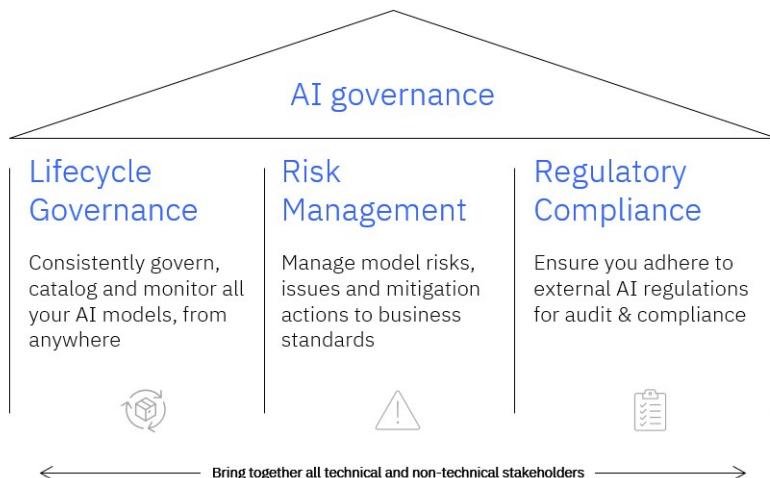
Name	Description	Model Owner	Model Class	Model Status	Tags
flan-t5-xl-3b Library > MRG > Foundation Models	flan-t5-xl (3B) is a 3 billion parameter model based on the Flan-T5 family. It is a pretrained T5 - an encoder-decoder model pre-trained on a mixture of supervised / unsupervised tasks converted into a text-to-text format, and fine-tuned on the Fine-tuned Language Net (FLAN) with instructions	Foundation Model	Proposed	<input type="checkbox"/>	
sam 2 Watsonx Innovation Studio		Prompt-based	Proposed	<input type="checkbox"/>	
sams insurance prompt Watsonx Innovation Studio		Prompt-based	Proposed	<input type="checkbox"/>	

Given the evaluation results, we may want to retrospective actions such as Issue Management or change the Risk level to Medium.

The screenshot shows the IBM Watsonx Governance console interface. A modal window titled "Use Case Data Gathering" is open, prompting the user to capture relevant information to this AI use case proposal and then submit using the Action button. The main form contains fields for General information (Name: Insurance claim summarisation – TEAM NAME, Status: Proposed, Description: GenAI Experience Workshop - Car Insurance Claim Use Case), Use Case Details (Additional Details), Risk (Risk Level: Medium), and Regulatory Information.

Congratulations! You have completed the Workshop. You have built, deployed and managed an AI Use Case tackling the 3 pillars of AI Governance.

Three pillars of AI governance





Optional steps: Navigating the UI of the Governance Console

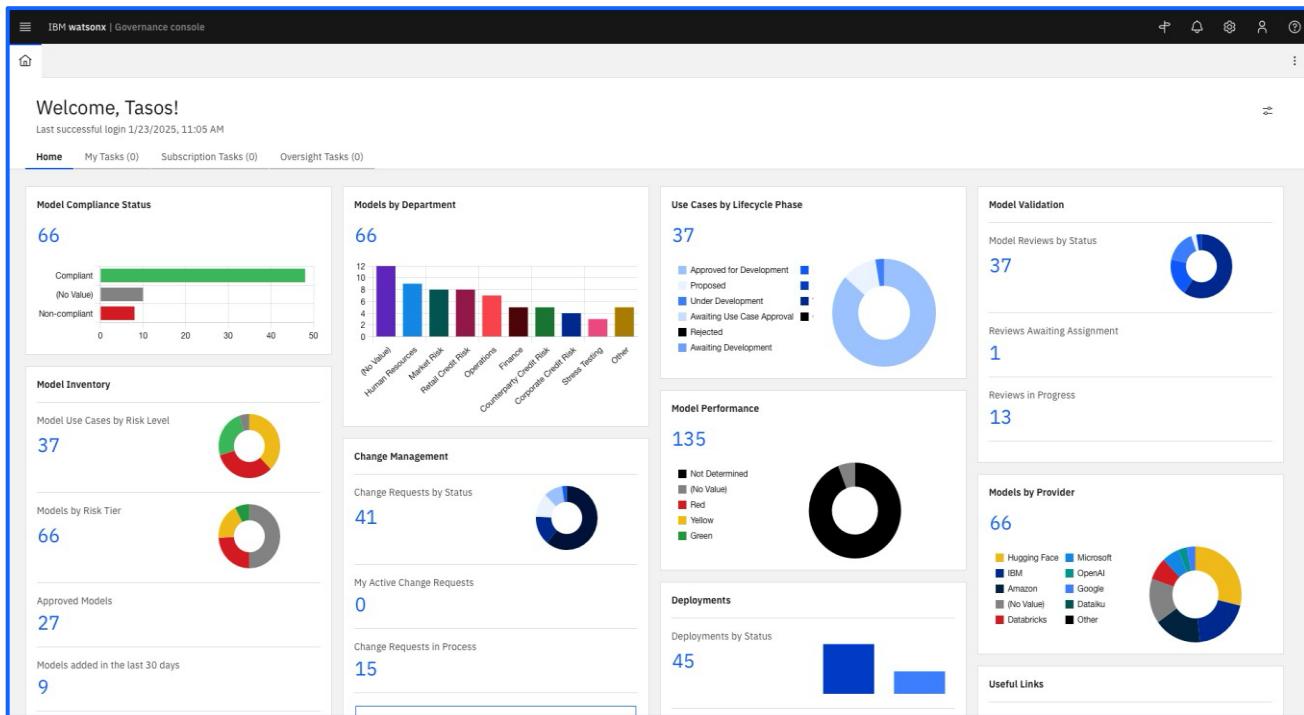
Get Started with your watsonx.governance
Workshop

1. Open **Chrome** or **FireFox** and navigate to the following [this](#) link
2. **Logon** with the credentials provided to you by the instructor

If an OTP is required, please ask your table facilitator to retrieve this.

Hint: make sure all spaces are removed from start and end of UN & Password

watsonx.governance User Interface



After login, you will be displayed the Homepage tab, and in it we will see four tabs.

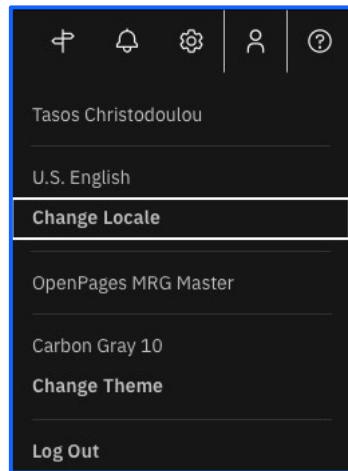
- **HOME** tab. The dashboard displays panels with content that is personalized for each user, intended to be relevant information for the user. *You may move and arrange them, as per your preferences. Later we will see how to create a new panel.*

The next three Task related tabs, are displayed tasks that I am notified during the workflow lifecycle of the solution components (*Click on each tab to see the content*):

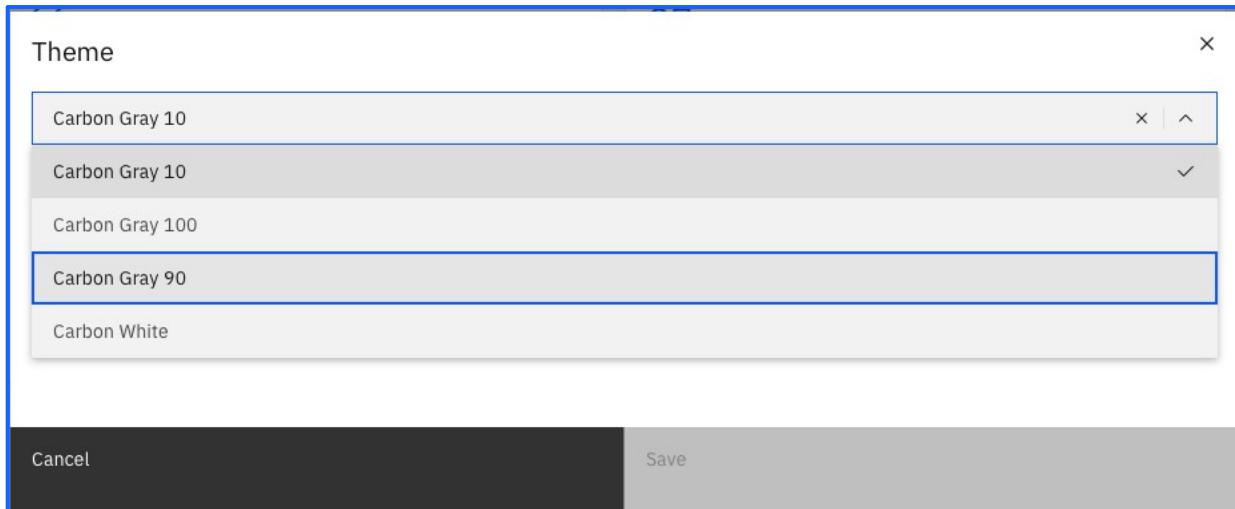
- **My Tasks:** activities that I have been assigned. It is action that I must perform.
- **Subscription Tasks:** activities that I want to know that happen, however I do not have to do anything.
- **Oversight Tasks:** I have full visibility on the whole lifecycle of an element (i.e., a Model Use case). I do not have any task assigned, just visibility.

Changing Colour Theme

1. Themes are used to customize the colours that are displayed in OpenPages (watsonx.governance).
2. Users can apply a theme by using the > Change Theme task.



3. Using the Drop-down list, change your Theme to Carbon Gray 90.

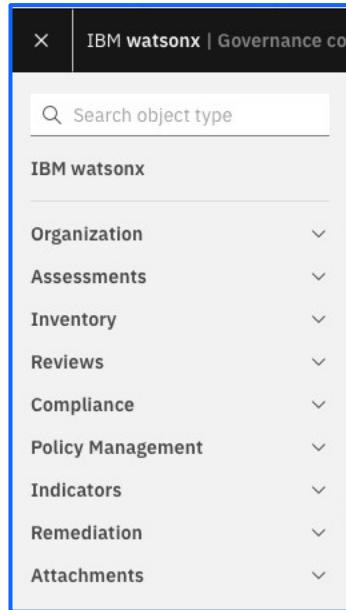


If you do not like the contrast, you may choose any other.

Themes can be configured by the client to allow a corporate colouring of the solution.

The Primary Menu

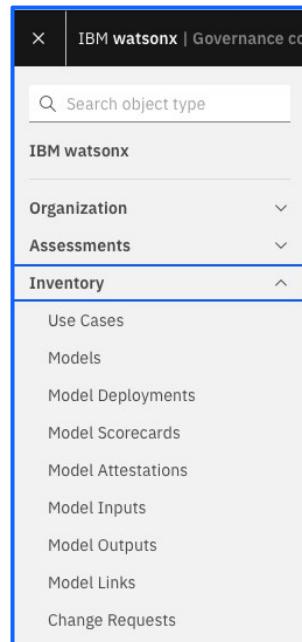
1. On the top left corner, click on the Primary Menu  (also known as the “Hamburger menu”)



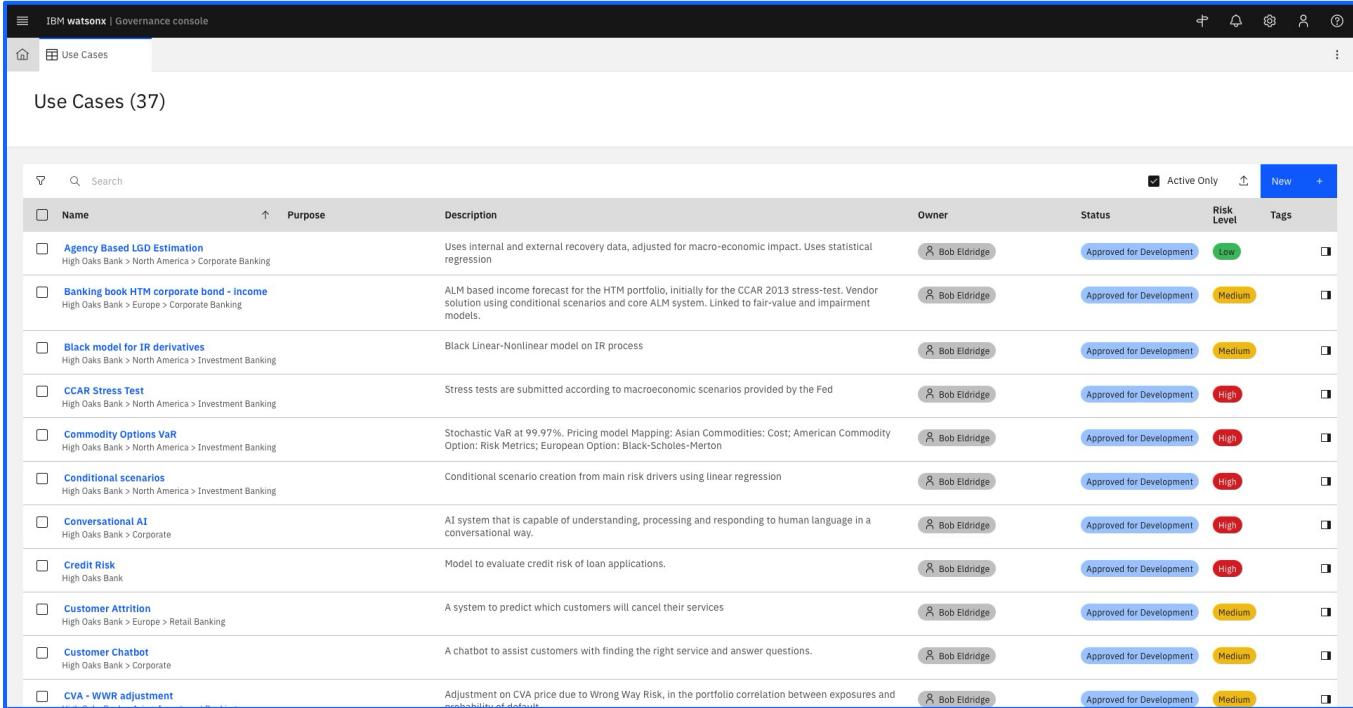
2. The Primary menu lists the categories of object types that you can access in IBM OpenPages® (watsonx.governance).

Expand and collapse the categories to see the object types. Seeing more or less objects depends on your profile.

3. When you click an object type, for example, **Use Cases** under the section Inventory, a list of objects opens.



4. Click on , expand **Inventory**, click on **Use Cases**.



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. Linked to fair-value and impairment models.	Bob Eldridge	Approved for Development	Medium	
Black model for IR derivatives	High Oaks Bank > North America > Investment Banking	Black Linear-Nonlinear model on IR process	Bob Eldridge	Approved for Development	Medium	
CCAR Stress Test	High Oaks Bank > North America > Investment Banking	Stress tests are submitted according to macroeconomic scenarios provided by the Fed	Bob Eldridge	Approved for Development	High	
Commodity Options VaR	High Oaks Bank > North America > Investment Banking	Stochastic VaR at 99.97%. Pricing model Mapping: Asian Commodities: Cost; American Commodity Option: Risk Metrics; European Option: Black-Scholes-Merton	Bob Eldridge	Approved for Development	High	
Conditional scenarios	High Oaks Bank > North America > Investment Banking	Conditional scenario creation from main risk drivers using linear regression	Bob Eldridge	Approved for Development	High	
Conversational AI	High Oaks Bank > Corporate	AI system that is capable of understanding, processing and responding to human language in a conversational way.	Bob Eldridge	Approved for Development	High	
Credit Risk	High Oaks Bank	Model to evaluate credit risk of loan applications.	Bob Eldridge	Approved for Development	High	
Customer Attrition	High Oaks Bank > Europe > Retail Banking	A system to predict which customers will cancel their services	Bob Eldridge	Approved for Development	Medium	
Customer Chatbot	High Oaks Bank > Corporate	A chatbot to assist customers with finding the right service and answer questions.	Bob Eldridge	Approved for Development	Medium	
CVA - WWR adjustment	High Oaks Bank	Adjustment on CVA price due to Wrong Way Risk, in the portfolio correlation between exposures and probability of default	Bob Eldridge	Approved for Development	Medium	

The list of Use Cases Display. This type of view is called Grid View. When you select an object type from the Primary menu, the **Home**, or a **Task View**, a Grid View is displayed.

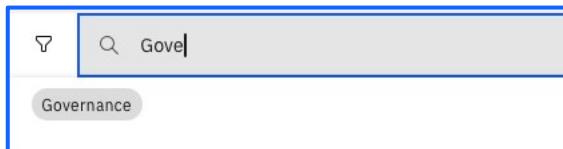
There are several actions that might be taken from a grid view.

Search Use Cases

1. Whilst in the Use Cases Tab, place the cursor on the Search box.



2. Start Typing “**Governance**”. You are presented with a Tag Suggestion



3. Select the Tag “**Governance**”.

Use Cases (2) Governance X

Notice the filter indicator displayed.

4. Place the cursor on the Search box once again.

▼ Search

5. Type “**Insurance**”. This is a Text Filter.

Use Cases (1) <small>Governance X</small>	
▼	<input type="text"/> Insurance
<input type="checkbox"/> Name	↑ Purpose
<input type="checkbox"/> Wx.demonstration Test	Watsonx Innovation Studio

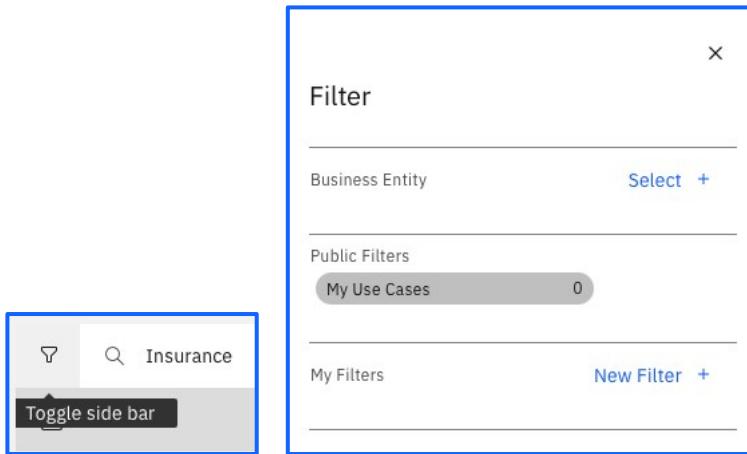
Notice that both filters have been applied.

6. Remove the filter display **Governance**.

Notice that now we have only a text search based on the word “insurance”.

These filters are **Ad hoc** and cannot be reused, as you must define them each time. Ad hoc filters are private filters that are not saved and only used in one session. Use an ad hoc filter if you do not frequently access the search criteria.

7. Click on the funnel symbol on the left to open the filter side bar.



Notice that we have a section named “Public Filters” and “My Filters” (private filters)

An administrator defines public filters. Users can customize private filters for their own use.

Use and customize private filters to find objects based on criteria on field values on objects and, optionally, parent objects. Users can customize private filters for their own use.

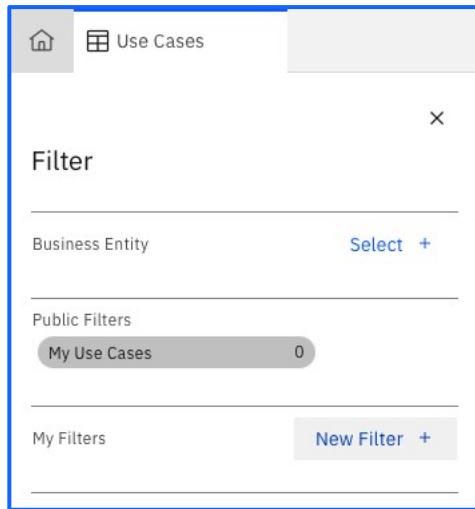
Only one public or private filter can be applied at a time.

You cannot simultaneously use a public filter and a private filter.

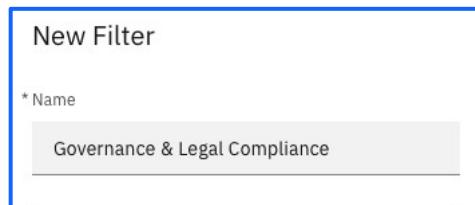
Let us create a filter that we may reuse.

Create Search Filters

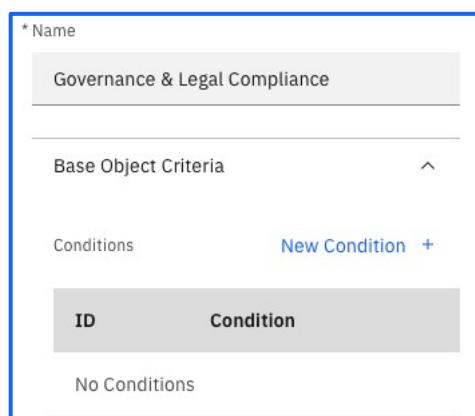
1. Click on **New Filter +**.



2. Provide the following **Name:** Governance & Legal Compliance



3. Click on **New Condition +**.



4. Select **Tag** & from the drop-down list select Governance & Legal Compliance.

New Condition

* Select a condition type.

Field Tag

* Tags *

2 × selected items ^

Data laws

Explainability

Fairness

Governance

Harmful code generation

5. Click **Save**.

Notice that now appears your filter under “My filters”.

Filter

Business Entity [Select +](#)

Public Filters [My Use Cases 0](#)

My Filters [New Filter +](#)

Governance & Legal Compliance 4

6. Your Filter is now displayed by the object type text.

7. Remove the filter and return to the homepage tab.

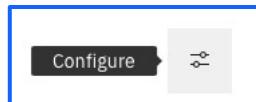
Use Cases (4) [Governance & Legal Compliance X](#)



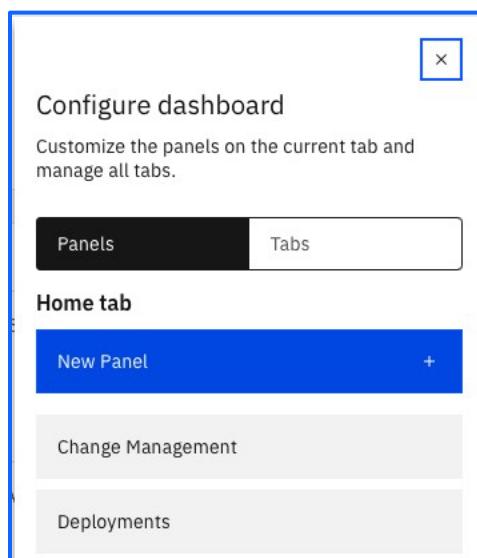
Configuring your home page: create a new panel

Now we are going to create a new panel, as we want to focus temporarily on Governance & Legal Compliance Use cases. We are going to design a panel that uses our just created private filter.

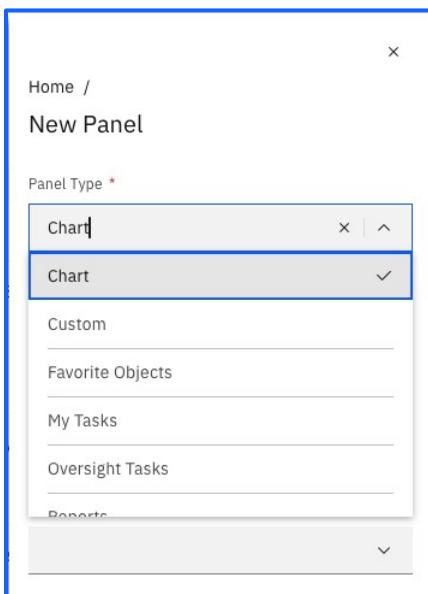
1. Click on the **Configure** icon on the right, above the tabs.



2. Select **New Panel +**.



3. From the drop-down list select **Chart**.



4. Accordingly provide the following parameters:

Label: Governance & Legal Compliance by Risk

Object Type: Use Case

Filter: Governance & Legal Compliance

Chart Type: Pie (you may also want to choose any other, except Gantt, to see how it looks like)

Chart Data Field: Risk Level

Method Type: Count

Home /

New Panel

* Label *

* Object Type *

* Filter

* Chart Type *

* Chart Data Field *

* Method Type *

Sort Direction

Cancel Done

5. Click Done & Save.

A new panel will appear in your dashboard.



6. Click on the categories in the legend to **hide/unhide** them.



7. Select any section on the chart to drill down on a Risk Level category and you will be taken to a grid view displaying your filter name and the filter based on the risk level you have clicked on.



Use Cases (1) Governance & Legal Compliance × Risk Level: High ×

Name	Purpose	Description	Owner	Status	Risk Level	Tags
Credit Risk High Oaks Bank		Model to evaluate credit risk of loan applications.	Bob Eldridge	Approved for Development	High	Legal compliance

8. If you wish to remove your panel from the homepage, click on the **Configure** icon on the right, above the tabs.



9. In the displayed list of panels, you would look for the one you would want to remove, hover the mouse over it and you would remove it by clicking on the minus sign.

Configure dashboard
Customize the panels on the current tab and manage all tabs.

Panels Tabs

Home tab

New Panel +

Change Management

Deployments

Governance & Legal Compliance by Risk

Issues by Status Delete

Model Compliance Status

Model Inventory

Model Performance

Model Validation

Models by Department

Models by Provider

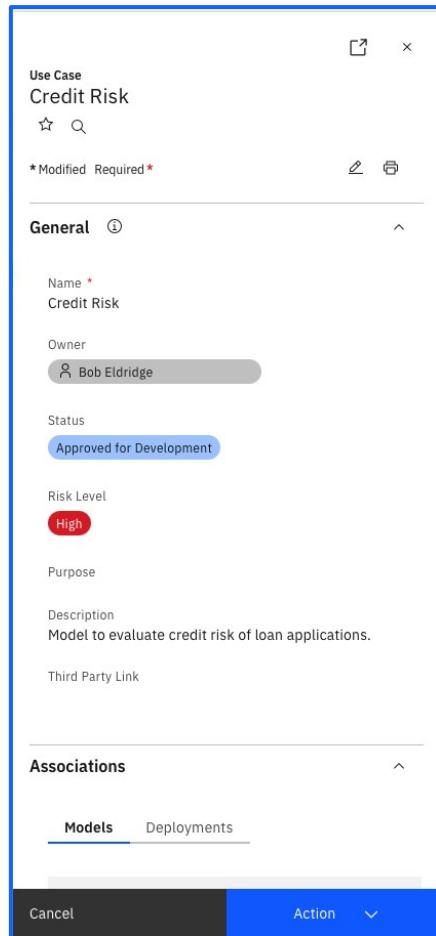
Do NOT remove the newly created panel as we will use it later on.

Navigation from an object

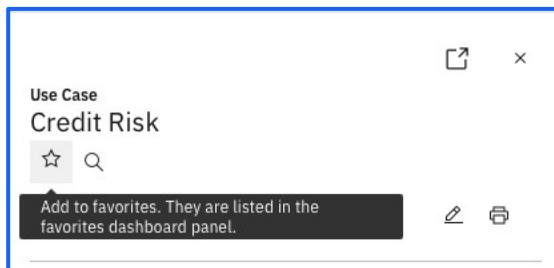
1. From the Homepage click on the “High Risk” level area in the chart of the panel you have just created.



2. Do a quick inspection of the record by clicking on the **Open Side Panel** on the right. 



3. Mark this Use Case as Favorite (it will make it appear in your “My Favorites” panel in your homepage), by clicking on the star icon to flag/unflag the record as Favorite.



4. Select the Use Case to open it in a new tab, as we want to investigate further.

The screenshot shows the 'Credit Risk' Use Case in the IBM Watson Governance console. Key details include:

- Name:** Credit Risk
- Status:** Approved for Development
- Risk Level:** High
- Description:** Model to evaluate credit risk of loan applications.
- Owner:** Bob Eldridge
- Purpose:** Model to evaluate credit risk of loan applications.
- Associations:** A table showing related models. One entry is visible:

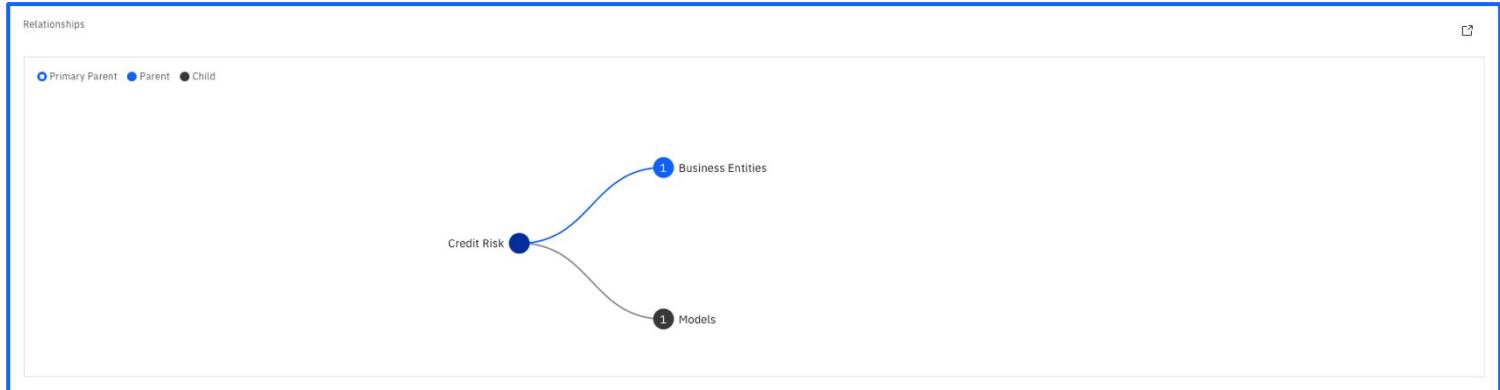
Name	Description	Model Owner	Model Status	Third Party Link	Tags
MOD-00001 High Oaks Bank	Credit Risk	Mark Owens	Pre Implementation Review		

5. Scroll down to the section **Related Models**. The Tab **All Models** displays the Models related to this Use Case.

The screenshot shows the 'Associations' section with the 'Models' tab selected. It displays a table with one entry:

Name	Description	Model Owner	Model Status	Third Party Link	Tags
MOD-00001 High Oaks Bank	Credit Risk	Mark Owens	Pre Implementation Review		

6. Scroll down to **Other Relationships**. The relationships tree view is displayed.



7. Open the tree diagram in a new tab by clicking on the **View fullscreen** icon
8. Following the picture below, click on the circles to expand the content. Click on the individual item to drill down on it. Notice the blue area that explains the relationships between the objects.



Audit Trail

1. Click back into the tab for the Credit Risk Use Case

The screenshot shows the IBM Watsonx Governance console interface. At the top, there is a navigation bar with icons for Home, Use Cases, and two instances of 'Credit Risk'. Below the navigation bar, the main content area is titled 'Use Case' and displays 'Credit Risk' with a star and a downward arrow icon. A blue border highlights the entire content area.

2. Switch from the **Task** tab to **Activity** tab.

The screenshot shows the 'Credit Risk' use case page. At the top, it displays 'Use Case' and 'Credit Risk' with a star and a downward arrow icon. Below this, there is a navigation bar with tabs: 'Task' (which is highlighted), 'Activity', 'Admin', and 'Security Performance Monitoring'. A blue border highlights the entire content area.

3. You may see all the changes made to this element. The details cannot be edited.

The screenshot shows the 'Activity' tab for the 'Credit Risk' use case. At the top, it displays 'Use Case' and 'Credit Risk' with a star and a downward arrow icon. Below this, there is a navigation bar with tabs: 'Task' (which is highlighted), 'Activity' (which is selected), 'Admin', and 'Security Performance Monitoring'. In the center, there is a table showing activity logs. The table has columns: 'Activity type', 'Old value', 'New value', 'Modified by', and 'Date and time'. Two entries are listed:

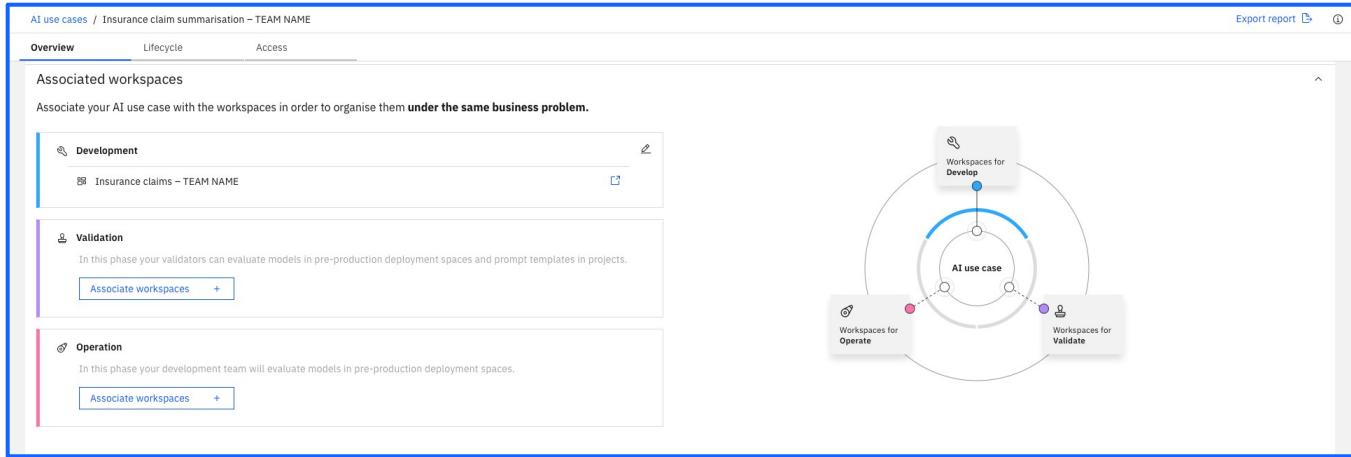
Activity type	Old value	New value	Modified by	Date and time
Tag - Added		Legal compliance	Tasos Christodoulou	Jan 24, 2025 6:38:44 AM EST
Model - Added		MOD-00001	System Administrator	Dec 5, 2024 1:48:43 AM EST

Promoting your prompt to Validation

N.B. This stage is skipped to save time. In real life, enterprises would you validation as a testing envrionemtn.

In between development and production their is a validation step. In this next section, we will walk through how to promote and validate your prompt through the stages of the model lifecycle.

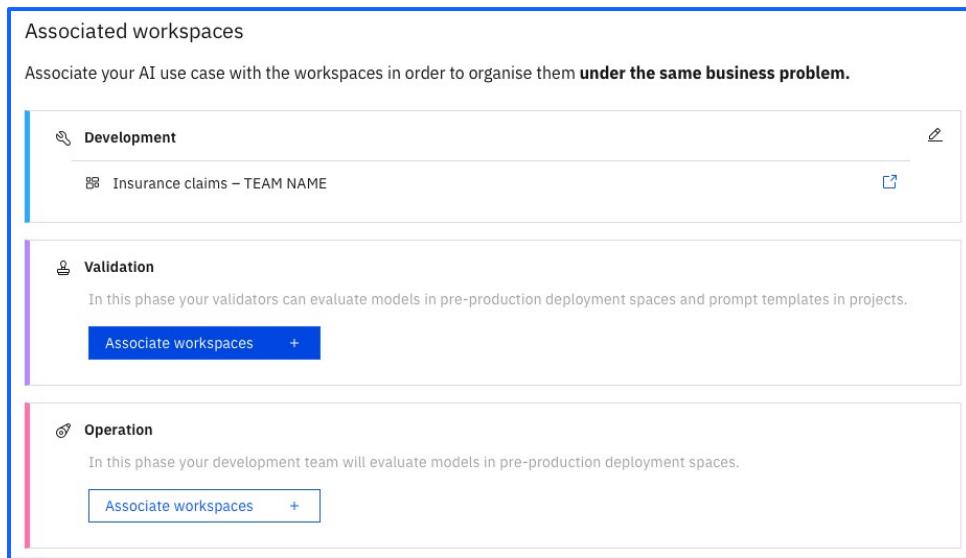
To start, navigate back to the **Overview** tab.



The screenshot shows the 'AI use cases / Insurance claim summarisation - TEAM NAME' overview page. The 'Validation' section is highlighted, showing a sub-section for 'Associate workspaces'. To the right, a circular diagram illustrates the model lifecycle stages: 'Workspaces for Develop' (top), 'AI use case' (center), 'Workspaces for Operate' (bottom-left), and 'Workspaces for Validate' (bottom-right).

Since we now want to promote the second prompt template to the validate phase, we will create a new deployment space and associate it to the **Validation** stage.

Click on the [Associate workspace +](#) button under the Validation space.

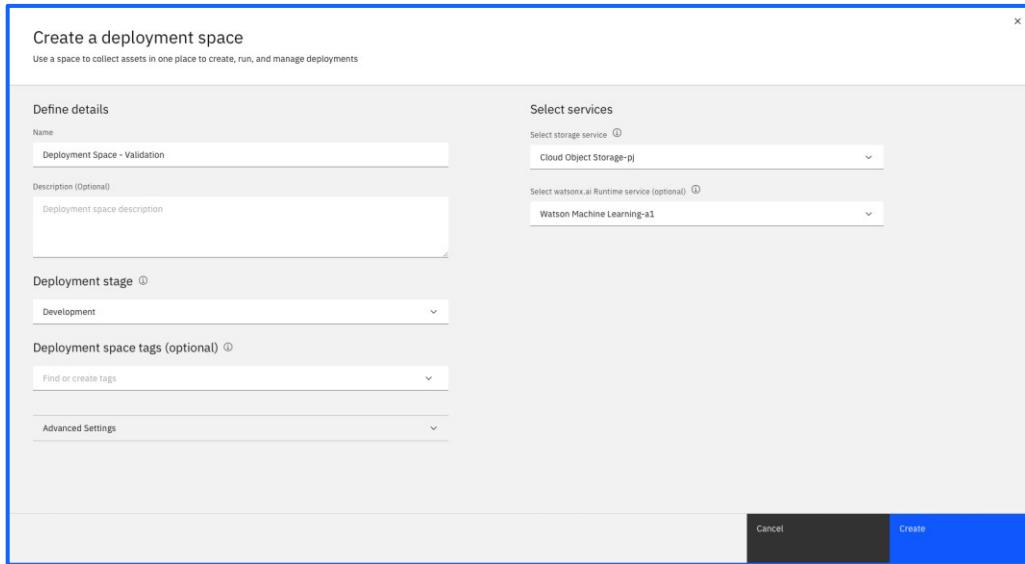


The screenshot shows the 'Associated workspaces' section of the overview page. The 'Validation' section is active, and the 'Associate workspaces +' button is highlighted with a blue border.

We will create a new **Deployment Space** for validation, to promote your prompt into.

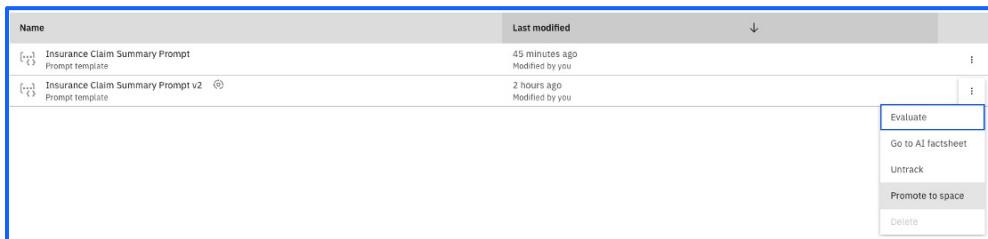
Give the **deployment space** a unique name as appropriate to this **validation stage** (i.e Deployment Space - Validation), so we can distinguish from other **deployment space** being created.

Select **Testing** as the deployment stage, as well as the default storage service and machine learning service as you did in the earlier section, before clicking [Create](#).

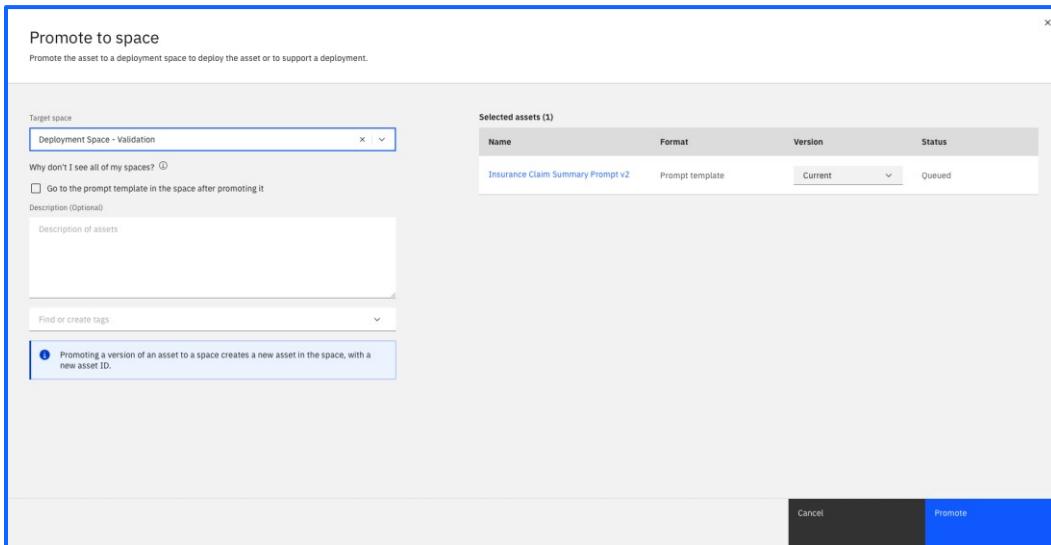


Associate this new Deployment Space under the Space section then click [Save](#).

Once this is completed, navigate back to your Project, and press the menu for the most recently created prompt template, and click Promote to space



Select the newly created deployment space under the **Target space** dropdown menu, and once completed click the [Promote](#) button.



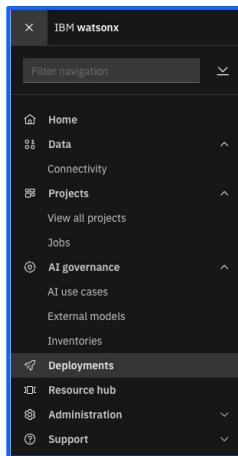
Navigate to the **Lifecycle** tab in your AI Use Case and note that it should appear under the Validate stage as opposed to the Develop stage of the AI use case lifecycle.

The screenshot shows the 'Default approach' tab of an AI Use Case. It has three main stages: Development, Validation, and Operation. The Development stage is described as 'Experiment with different approaches of ML development or prompt engineering and test them in projects.' The Validation stage is described as 'Evaluate models and prompt templates with validation data.' The Operation stage is described as 'Evaluate and monitor models and prompt templates in production.' Below these stages, there are two versions listed: 'v0.1.0' and 'v0.0.1'. Each version has a tree view showing its components. 'v0.1.0' includes 'Insurance claims - TEAM NAME' and 'Deployment Space - Validation'. 'v0.0.1' includes 'Insurance claims - TEAM NAME'. To the right of each component tree, there are status messages: 'No models promoted to a production space.' for both versions.

The prompt has not been deployed yet – so this is what we will be doing in the next section.

Deploying the prompt

Use the navigation menu to move into the **Deployments** section.



Click into the relevant validation space (i.e. the space you created in the previous section) within the **Spaces** tab.

The screenshot shows the 'Deployment Spaces' tab. It displays a single deployment space named 'Deployment Space - Validation'. The space was last modified on Jan 28, 2025 at 5:19 PM by an Admin. The space is categorized under 'Development' and has 0 online deployments and 0 jobs. A 'New deployment space' button is located in the top right corner.

Move into the on the **Assets** tab, and click the three dots to the right of the most recent prompt template as shown:

The screenshot shows the 'Assets' tab for the 'Deployment Space - Validation'. It lists one asset named 'Insurance Claim Summary Prompt v2', which is a 'Prompt template'. The asset was last modified 7 minutes ago by a Service user. There are buttons for 'Import assets' and 'Options' on the right side of the asset row.

Finally, click the option to Deploy the prompt.

The screenshot shows the 'Deployment Space - Validation' interface. At the top, there are tabs for 'Overview', 'Assets' (which is selected), 'Deployments', 'Jobs', and 'Manage'. Below the tabs, there is a search bar labeled 'Find assets' and a button 'Import assets'. On the left, there are filters for '1 asset' (with 'All assets' selected) and 'Asset types' (with 'Prompts' selected). The main area displays a table with one row for 'Insurance Claim Summary Prompt v2'. The columns in the table are 'Name' and 'Last modified'. The asset details show '7 minutes ago' and 'Service'. To the right of the table, there are buttons for 'Deploy' and 'Delete'.

Within the deployment wizard, give an appropriate name (i.e. Insurance Summarisation - Deployment) before clicking [Create](#).

The screenshot shows the 'Create a deployment' wizard. The first step is 'Define details'. It includes fields for 'Associated asset' (set to 'Insurance Claim Summary Prompt v2'), 'Deployment type' (set to 'Online'), 'Name' (set to 'Insurance Summarisation - Deployment'), 'Serving name' (set to 'Deployment serving name'), 'Description' (set to 'Deployment description'), and 'Tags' (set to 'Add tags to make assets easier to find'). At the bottom right, there are 'Cancel' and 'Create' buttons.

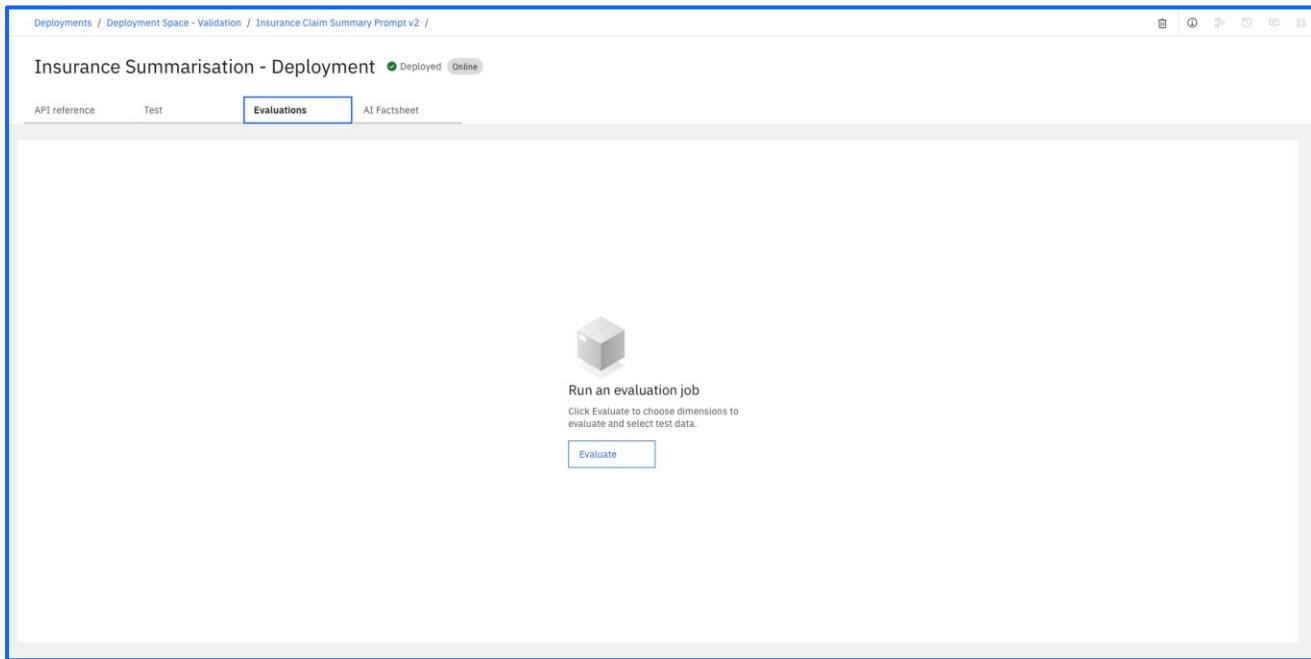
If you navigate back to the AI use case, you will see the prompt as Pending Evaluation.

The screenshot shows the AI use case interface. It displays a 'Default approach' section with 'Latest: 0.1.0 | 2 versions'. Below it, there is a summary: 'A default approach for tracking your AI assets.' Under the 'Development' section, there is an experiment titled 'Experiment with different approaches of ML development or prompt engineering and test them in projects.' Under the 'Validation' section, it says 'Evaluate models and prompt templates with validation data.' Under the 'Operation' section, it says 'Evaluate and monitor models and prompt templates in production.' At the bottom, there is a tree view showing 'v0.1.0' which contains 'Insurance claims - TEAM NAME' and 'Deployment Space - Validation'. 'Deployment Space - Validation' contains 'Insurance Claim Summary Prompt v2' and 'Insurance Summarisation - Deployment' (which is marked as 'Pending Evaluation').

This is the stage at which an appropriate stakeholder would evaluate and start monitoring the prompt – which we will be doing in the next and final section.

Activate prompt monitoring (Evaluation)

Click on the prompt that is pending evaluation, and move into the **Evaluations** tab to be taken to the screen shown below:



Click on the [Evaluate](#) button in the centre of the screen to start another evaluation before checking the results. It may seem redundant to run evaluations with the same prompt and the same data set over and over, but it is important to remember:

- The evaluations would be performed by different people in different roles, e.g. the prompt developer would not be the one evaluating it for use in production.
- In a real-world scenario, the data sets would differ between evaluations.
- There is a strong likelihood that the different models will change the evaluation scores.

After clicking [Evaluate](#), navigate back to the AI use case to confirm that the status has changed from Pending to **Evaluated**

^  **Default approach**
Latest: 0.1.0 | 2 versions

A default approach for tracking your AI assets.

⋮ →

Development	Validation	Operation
Experiment with different approaches of ML development or prompt engineering and test them in projects.	Evaluate models and prompt templates with validation data.	Evaluate and monitor models and prompt templates in production.

◇ v0.1.0

Development	Validation	Operation
 Insurance claims – TEAM NAME  Insurance Claim Summary Prompt v2 ⓘ	 Deployment Space - Validation  Insurance Claim Summary Prompt v2 ⓘ  Insurance Summarisation - Deployment ⓘ Evaluated	No models promoted to a production space.