AI & Automation Unpacked Hackathon

{hack}

IBM.

AI & Automation Unpacked Hackathon

Contents

The hackathon challenge

In this hackathon, participants are expected to leverage the powerful **IBM Granite model(s)** to design and build a **proof-of-concept AI-powered automation solution** that streamlines everyday business processes, enhances efficiency, and drives industry transformation. You and your team should determine the specific problem or use case you want to address, and use what you've learned to develop your big idea using IBM Granite to solve it.

Participants can access IBM Granite models through either the open-source or the IBM watsonx.ai platforms. Instructions for accessing and using Granite models from both sources are provided in this guide.

IBM Granite is a family of AI models purpose-built for business, engineered from the ground up to ensure trust and scalability in AI-driven applications. These enterprise-ready models deliver exceptional performance against safety benchmarks and across a wide range of enterprise tasks, from cybersecurity to retrieval-augmented generation (RAG). By leveraging the power of IBM Granite models, businesses can automate complex processes, enhance decision-making, and drive innovation across various domains.

The latest **IBM Granite 3.3** represents a significant expansion of IBM Granite's multimodal capabilities. Headlined by **Granite Speech 8B**, IBM's first official speech-to-text model, Granite 3.3 excels in automatic speech recognition (ASR) and automatic speech translation (AST). This model is built on top of **Granite 3.3 8B Instruct**, which offers enhanced reasoning capabilities, including fill-in-the-middle (FIM) text generation. These advancements make Granite 3.3 ideal for automating tasks that involve audio processing, complex reasoning, and real-time data retrieval. Learn more about Granite 3.3.

Looking ahead, **IBM Granite 4.0** introduces the **Granite 4.0 Tiny Preview**, a compact and compute-efficient model designed for long-context tasks and instruction-following scenarios. With its hybrid Mamba-2/Transformer architecture, Granite 4.0 Tiny delivers top-tier performance with significantly reduced memory requirements, making it suitable for running on consumer-grade hardware. This model is perfect for automating workflows that require handling extensive context, such as document analysis, dialogue summarization, and knowledge-intensive question answering. Learn more about Granite 4.0. Together, Granite 3.3 and Granite 4.0 empower businesses to build robust AI automation solutions that are both powerful and efficient.

Refer to example use cases to help you get started with your solution idea.

Note on data sets before you begin

Participants are required to bring their own datasets to build the solution aligning to your use case. As you collect data for your project, you'll want to use the best practices. Here are some helpful tips:

- Teams are responsible for ensuring data is compliant.
- Data from public websites may be used, if the terms allow for commercial use, but please keep a list of the websites you use.
- Do not use data or assets containing company confidential data, or any other data without permission from the data owner. Teams are responsible for getting approval.
- Do not use any client data.
- Do not use any data containing personal information (PI).
- Do not use data obtained from social media.

Get started with IBM Granite

Participants can use IBM Granite models to build their innovative AI solution by accessing them through the following options:

1. [Open-source platforms](#)

2. [IBM watsonx.ai](#)

Option 1: Open-source platforms

To access IBM Granite models through open-source platforms and run them locally, you have the following open-source options:

- [Download Granite on Hugging Face](#)
- [Run locally with Ollama](#)
    - You have to download and install Ollama on your local machine to use the IBM Granite models.
        - [Mac](#)
        - [Windows](#)
        - [Linux](#)
    - [Granite models supported on Ollama](#)

System requirements

To run Granite models locally, it is recommended to use a machine with **at least 32 GB RAM** and a **GPU processor**. While running the models on a lower size RAM and CPU is possible, it may result in **slower performance**.

IBM Granite documentation

Refer to [IBM Granite documentation](#) to explore all the [Granite models](#) and recipes for [agents](#) and other [use cases](#) to help you get started.

Granite cookbooks

Try sample use case of Granite cookbooks to understand and use different Granite models for your solution use case. Refer [Granite cookbooks page](#).

Granite Workshop

Try the Granite Workshop to get hands-on experience for a few use cases that demonstrate the value of generative AI using Granite models. By the end of this workshop, you will be able to:

- Summarize a text document using text summarization
- Generate specific information from a large document using the RAG technique
- Predict future trends using time series forecasting
- Generate programming code (Bash) by prompting a code model

BeeAI Agentic Framework

BeeAI is an open-source agentic framework for building production-ready **AI agents**. The framework is specifically optimized to help you build powerful AI agents with smaller models such as the Granite 3 series.

- Get started with your language of choice
  - BeeAI Framework for Python
    - Installation
    - Try quick example
  - BeeAI Framework for Typescript
    - Installation
    - Try quick example
  - Head to BeeAI GitHub for more documentation and examples.

- New to building AI agents? Try our hands-on labs to ramp up!
  - Get familiar with the basics such as PromptTemplates, Messages, Memory and how to setup and generate output using a ChatModel.
  - Try out the Granite-powered ReActAgent and connect it existing tools, or create your own.
  - [Advanced] Solve complex use cases by building an AI agent as a workflow. This low-level implementation gives you the most control and flexibility to define your single agent or multi-agent implementation.

Other Granite quick start samples

- Build an AI-powered multimodal RAG system with Docling and Granite
- Build an AI research agent for image analysis with Granite 3.2 Reasoning and Vision models
- Building data pipelines using natural language with Data Prep Kit (DPK)
- Build a local AI co-pilot using IBM Granite Code, Ollama, and Continue
- Build a multi-agent RAG system with Granite locally

Option 2: IBM `watsonx.ai`

To access and use IBM Granite models on the watsonx.ai SaaS runtime environment for model inferencing, participants must be registered for the hackathon and must request an IBM Cloud account. Follow the instructions below to request access to the watsonx.ai platform.

If you are working as part of a team, please plan to collaborate using one team member's cloud account. Adding or removing team members is not supported on the hackathon-provisioned IBM Cloud accounts.

Note on IBM Cloud service usage

For this hackathon, **$100 credits** will be automatically applied on the provisioned **IBM watsonx.ai platform**. This should be sufficient for designing and creating a compelling submission.

You will receive periodic email notifications about your **credit consumption** at the following usage levels: **25%, 50%,** and **80%**. Once you reach **100% usage**, your account will be **suspended**. You can appeal the suspension by completing the form shared in the account suspension notification email.

Please note that these email notifications are sent **once per hour**, so there is a possibility that you may **exhaust all your credits before receiving an alert**.

Please plan to use the watsonx.ai efficiently and back up your work accordingly. Refer **[tips to work efficiently on watsonx.ai platform](#)** (Tokens and CUH explained) and **[saving your work](#)**.

Important:

- **Foundation model inferencing** consumes tokens, which are measured as Resource Units (RUs).
  1,000 tokens = 1 RU, and each RU costs **$0.0001 USD.**
  [Learn more about tokens and tokenization](#).

- If you are using **Jupyter Notebook editor on watsonx.ai,** consider selecting **a lower runtime environment** to avoid high resource consumption and quickly depleting your credits. Notebook runtimes are billed based on **Capacity Unit Hours (CUH)** at a rate of **$1.02 USD per CUH.**
  [Learn more about capacity unit hours and watsonx.ai Studio pricing plans](#).

Note on available services

The IBM Cloud and the watsonx.ai platform are **pre-configured with only the services required** to complete the hackathon. If you notice a permission/access issue for any service or the cloud catalog, then they are not required/available for this hackathon.

These features/capabilities are out of scope for this hackathon:

- Agent Studio (Beta)

- Deploy on IBM Cloud/watsonx.ai (including Deployment space)
- Bring your own model
- Fine tuning models
- AutoAI pipeline
- SPSS Modeler
- Federated Learning
- Cloud Object Storage service

Important: You are required to use only the IBM Granite models on the watsonx.ai platform for this hackathon. The use of any other models is out of scope and may negatively impact the evaluation of your submission.

The hackathon provisioned IBM Cloud account will be deactivated after the completion of the hackathon. Please plan to [save your work](#) at the end of the hackathon.

Request your IBM Cloud account

Participants must request an IBM Cloud account to access and use IBM Granite models on the watsonx.ai platform. Follow the steps below to request your cloud account:

1. Access the hackathon's IBM Cloud account request link on your browser - https://www.ibm.com/account/reg/us-en/signup?formid=urx-53809.

2. Create an IBMid for your hackathon registered email by completing the account information and click "**Next**" button. If you already have an IBMid, proceed to log in, complete the authentication process and skip to step 4.

3. Verify your email by entering the 7-digit code sent to your email and select the "**Submit**" button.

4. Complete the IBM Cloud account request by verifying your name, and email and select the "**Request account**" button.

5. Once your request is successfully submitted, you will receive an email invite to join the cloud account, and you will be navigated to the IBM TechXchange Dev Day site. Next, follow the "Access your IBM Cloud account" instructions to access your hackathon IBM Cloud account.

Access your IBM Cloud account

After requesting a cloud account, participants will receive an email invite to join the cloud account. Follow the steps below to access your cloud account:

1. Check the email inbox you used to register for the hackathon and open the email you received from the IBM Cloud team about joining your cloud account. Please check your junk/spam folders if you are not able to find the email in your inbox. You can also quickly search for "IBM Cloud" to locate the email.
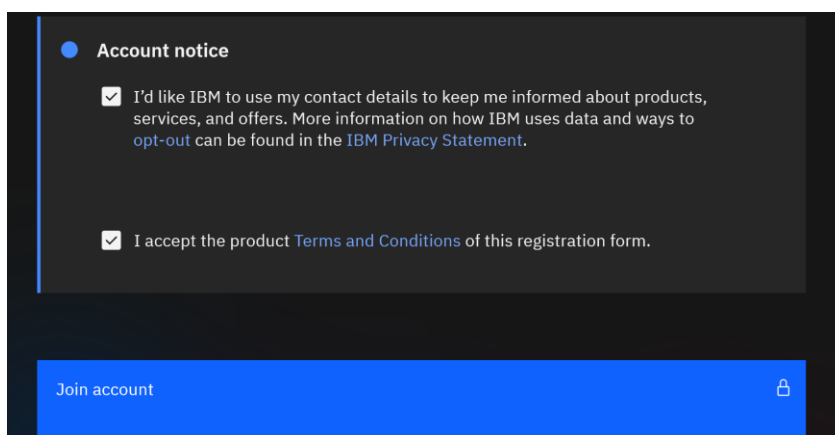
2. Click the **Join Now** button in that email. A new browser tab will open with the cloud account sign up page.
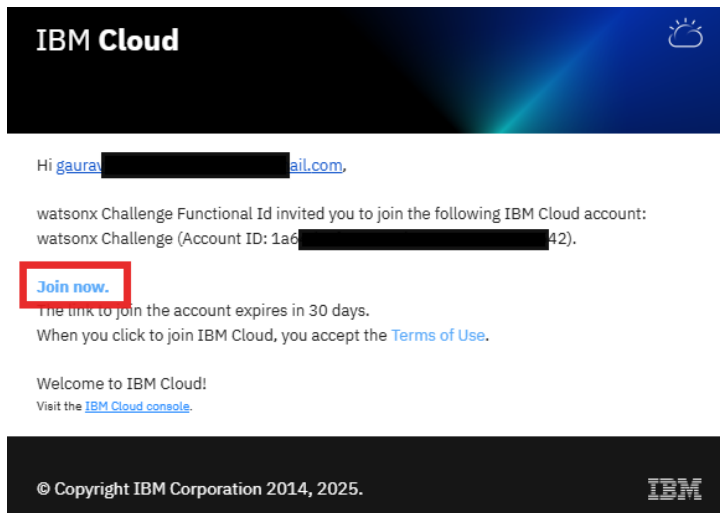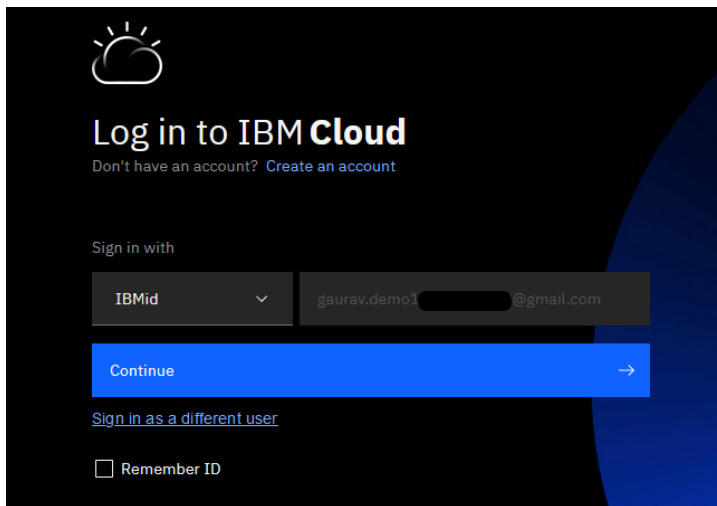

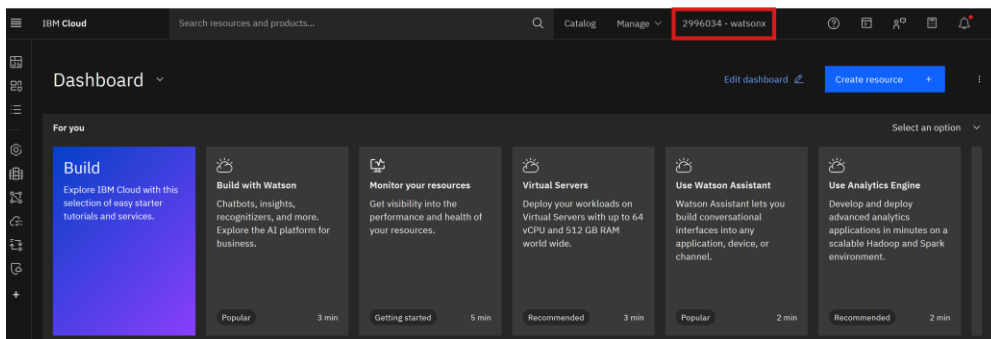3. Review your account and personal information.
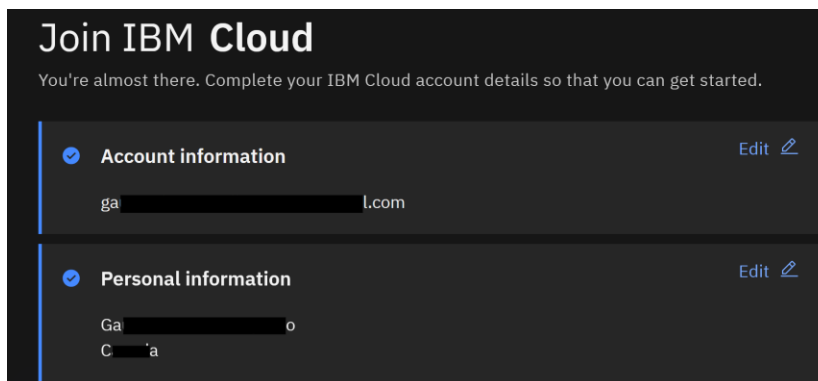
4. Read and accept the Account notice and click the **Join Account** button.

5. Complete the authentication process by clicking the **Continue** button. You may be prompted to enter your password.

6. After you authenticate successfully, you will be taken to the IBM Cloud dashboard.

7. If you have an existing personal IBM Cloud account for the same email/IBMid, sometimes you will be directed to your personal account. In this case, please switch your account to the **xxxxxxx - watsonx** account. Select your account drop-down at top-right of the dashboard and select watsonx account. Refer to the below image on switching accounts in your cloud dashboard.
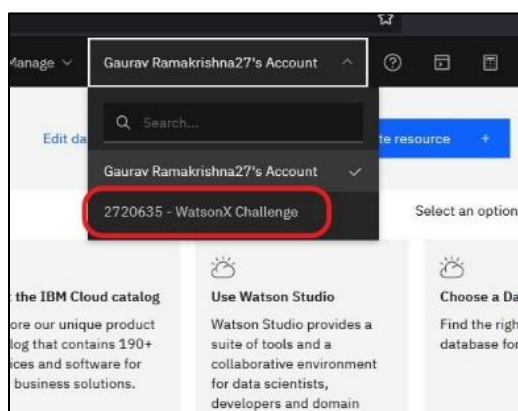
Access Prompt Lab on watsonx.ai platform

After successfully joining the IBM Cloud account, you can now access the Prompt Lab on watsonx.ai platform to work with the Granite models supported on the platform and build your solution.

1. Log in to the watsonx.ai platform (https://dataplatform.cloud.ibm.com/wx/home?context=wx) with the email you used to access your IBM Cloud account. If you are asked to select a region, select the "**Dallas**" region.

    Important: If you see the error message "**Insufficient permissions to provision instances**", close the current browser tab, open a new one, and try accessing the watsonx.ai platform again.
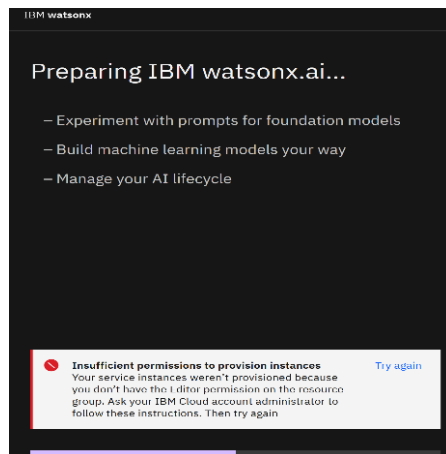

2. After successful authentication, you will see "Welcome to watsonx". You can either take the tour or skip it.

3. Next, you will see the watsonx.ai dashboard. Ensure the name of the account is "**xxxxxxx – watsonx**" and the region is "**Dallas**".
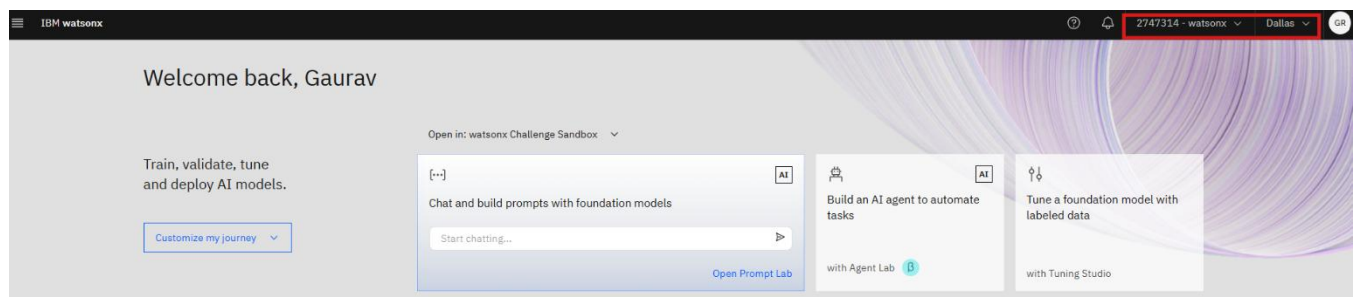
4. Select the "**Open Prompt Lab**" button on the "Chat and build prompts with foundational models" widget.
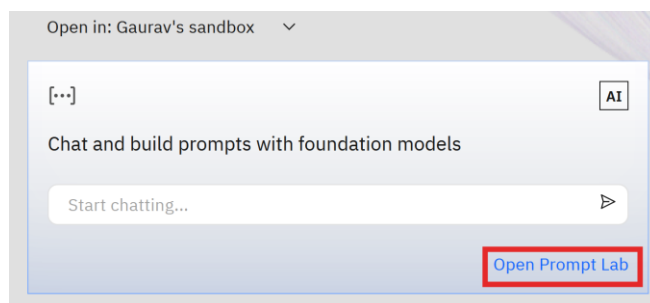
5. The "Welcome to Prompt Lab" tour will be displayed. You can take the tour to get a quick introduction or skip it.

6. The Prompt Lab Editor opens with a chat window to get you started with the prompt session.
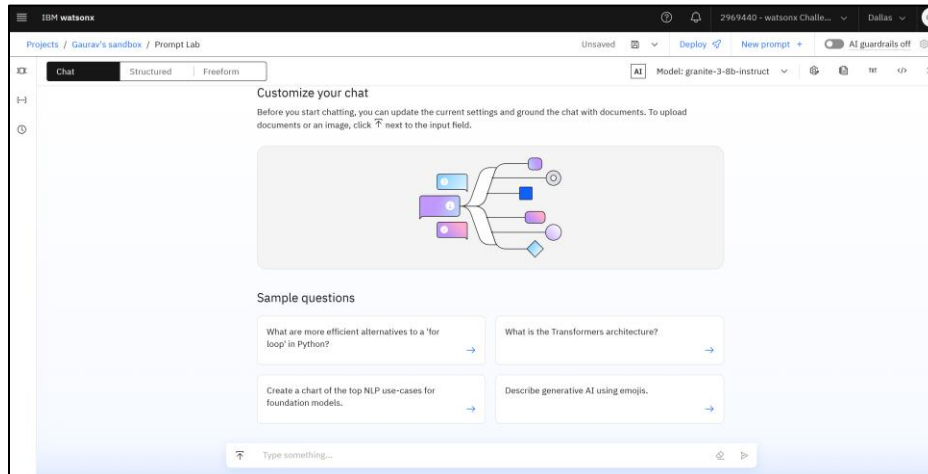
Work with the watsonx.ai Prompt Lab

The watsonx.ai Prompt Lab is an easy-to-use prompt engineering interface where you can experiment prompting different IBM Granite foundation models, explore sample prompts, tune model parameters, integrate applications with an API endpoint, and save and share your best prompts.

Take a tour of the Prompt Lab and try the interactive demo.

You can access and use the **IBM Granite models** to build your innovative solution using Prompt Lab.

Prompt Lab editor

In the Prompt Lab, you can experiment with prompting different foundation models, explore sample prompts, as well as save and share your best prompts. The Prompt Lab editor is a great place to experiment and iterate with your prompts. Try the quick start lab.

However, you can also prompt foundation models in watsonx.ai programmatically. Refer to "Programmatic access (API/SDK)" section.

Selecting an IBM Granite model

The "**llama-3-3-70b-instruct**" model will be pre-selected by default in the Prompt Lab editor. You will need to change it to a Granite model. To select a Granite model:

1. Select the AI Model drop-down menu at the top-right of the editor and select "**View all foundation models**".
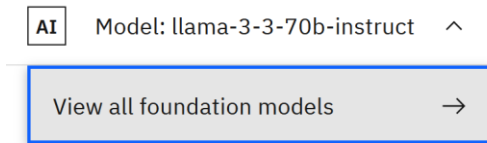
2. The "**Select a foundation model**" widget will appear. Enter "granite" in the search bar. All the granite series models will be displayed. You can select a model tile to learn about the model and use it.

   Important: You are required to use **ONLY IBM Granite models** exclusively on watsonx.ai. Using any other models will negatively impact the judgement of your submission and are out of scope for this hackathon.
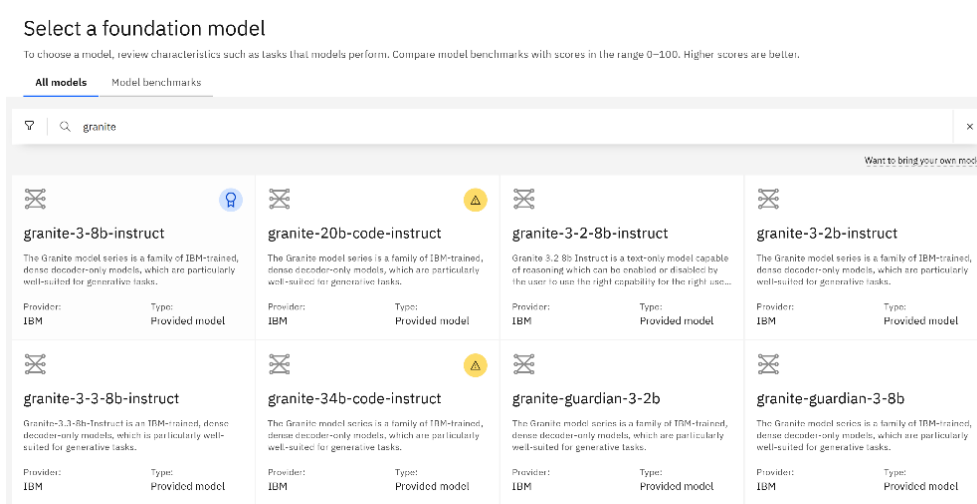
   To understand how models can address your use case, including information on model modalities, supported languages, tuning, and indemnification, see our product documentation on choosing a model.

   Note: Bigger models are not always better. Learn why smaller models can be better and more cost effective.
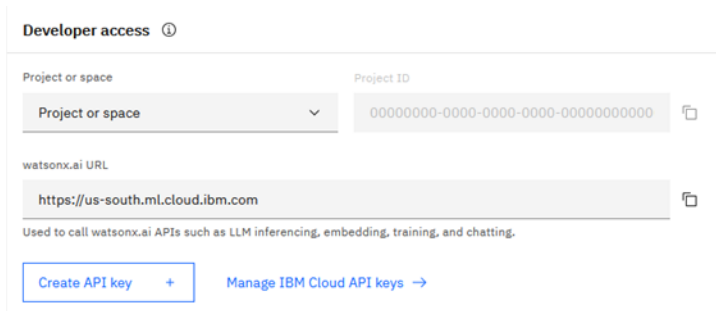
Programmatic access (API/SDK)

You can inference the watsonx.ai models with API or SDK requests.

Developer access information

To use the supported watsonx.ai APIs/SDKs, you will need three values: a **project ID**, an **endpoint URL** and an **API key**.

- Go to watsonx.ai home page.

- Scroll down to the "**Developer access**" section.
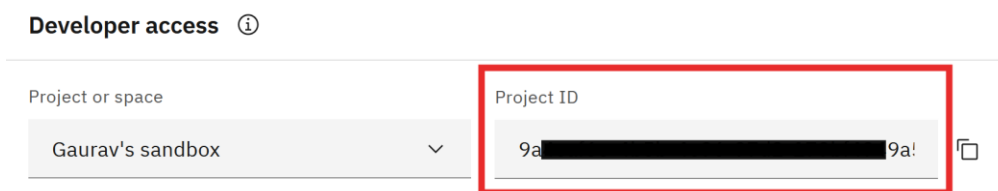
- Select the "**Project or space**" drop-down and select your sandbox project option. A **project ID** will be displayed.
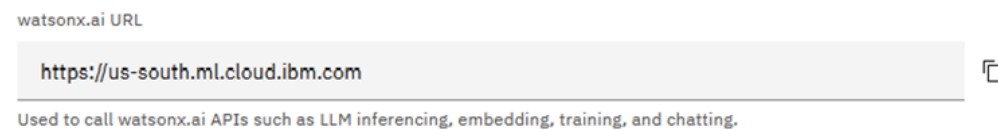 **Note**: A space ID is **not required** as it is out of scope for the hackathon.
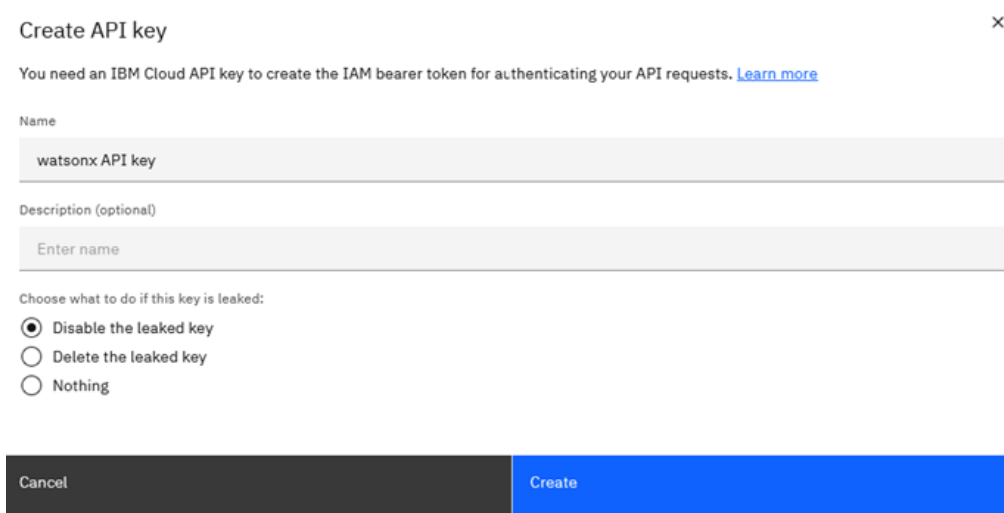

- A default **watsonx.ai endpoint URL** will be displayed for the Dallas region. Ensure the region is always set to **Dallas** at the top right of the watsonx.ai home page.

- Select the "**Create API key**" button. A **Create API key** widget will be displayed. Enter a name, provide optional description and choose the "Disable the leaked key" option. Click the "**Create**" button.

- An API key will be created successfully. Copy the API key and save it safely to use for calling the API/SDK. You can also download and save the file in a secure path in your system.
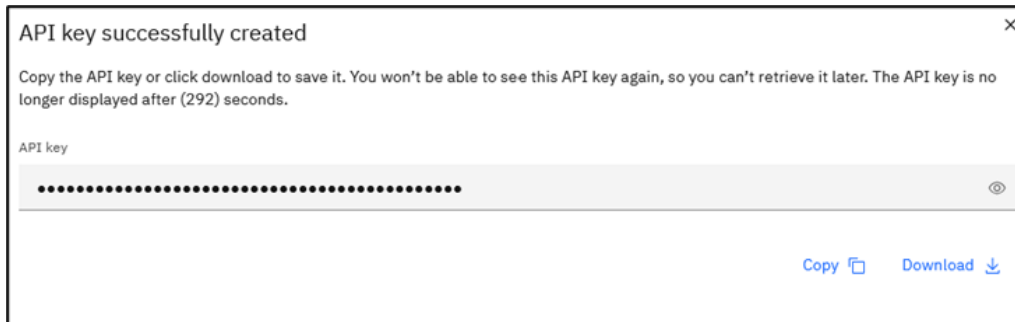
watsonx.ai programmatic options

There are multiple options to help you get started using watsonx.ai APIs/SDKs.

Option 1: Prompt Code on Prompt Lab

Refer to the access prompt code instructions to learn how to quickly get access to the text generation API within the watsonx.ai Prompt Lab.

Option 2: Different watsonx.ai API capabilities

Explore and leverage different watsonx.ai API capabilities in your solution.

- Chat
- Tool calling
- Text generation
- Time series
- Text rerank
- Embeddings
- Text extraction

Refer supported API functionality by model here.
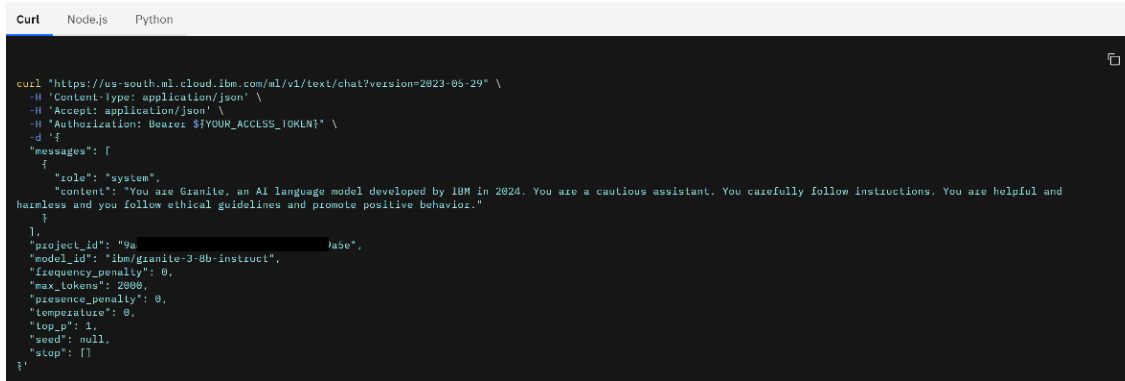
Access the prompt code (API) from Prompt Lab editor

To prompt an AI model programmatically, you can view and copy the prompt code by selecting the **View code** icon </> at the top-right of the prompt lab editor.

The prompt code is available as a Curl, Node.js and Python.

You will require an IAM access token to authorize the prompt code and need to replace **${YOUR_ACCESS_TOKEN}** placeholder in the prompt code. You can create an IAM access token using an API key.

- **API key:**

  Refer to [Developer access information](#) to get an API key.

- **Generate IAM Access Token:**

  Programmatically generate an IAM access token with the API key using the following cURL command:

```
curl -X POST 'https://iam.cloud.ibm.com/identity/token' -H 'Content-Type: application/x-www-form-urlencoded' -d 'grant_type=urn:ibm:params:oauth:grant-type:apikey&apikey=MY_APIKEY'
```

- **curl -X POST** → Specifies an HTTP **POST** request.
- **URL (https://iam.cloud.ibm.com/identity/token)** → The endpoint to request an authentication token from IBM Cloud.
- **-H "Content-Type: application/x-www-form-urlencoded"** → Sets the request header to indicate that the data is sent in form-encoded format.
- **-d (Data Payload)** → Sends the required data:
- **grant_type=urn:ibm:params:oauth:grant-type:apikey** → Specifies the OAuth grant type as API Key.
- **apikey=MY_IBM_CLOUD_API_KEY** → Replace MY_IBM_CLOUD_API_KEY with your actual IBM Cloud API key.

```
Expected Response:

{
    "access_token": "eyJhbGciOiJIUz......sgrKIi8hdFs",
    "refresh_token": "not_supported",
    "ims_user_id": 13……55,
    "token_type": "Bearer",
    "expires_in": 3600,
    "expiration": 1473188353,
    "scope": "ibm openid"
}
```

Note: An IAM token is valid for up to 60 minutes, and it is subject to change. When a token expires, you must generate a new one. Use the property "*expires_in*" for the expiration of the IAM token that you have just created.

Quick start hands-on exercises

Try the quick start exercises and notebooks for sample use cases to get started with using watsonx.ai.

Important notes:

- Refer to developer access information section to use watsonx.ai credentials as you try the exercises.
- Some of the exercises could include the usage of old Granite model version. You can replace them with newer versions for better performance and output. To check the latest supported Granite models on watsonx.ai, either follow selecting an IBM Granite model on Prompt Lab or refer to supported Granite models on watsonx.ai.
- The hackathon provisioned cloud accounts **do not support solution deployment.** You can run your solution deployment locally on your machine and showcase them in your submissions.
- **Foundation model inferencing** consumes tokens, which are measured as Resource Units (RUs). **1,000 tokens = 1 RU,** and each RU costs **$0.0001 USD.**
  Learn more about tokens and tokenization.
- If you are using **Jupyter Notebook editor on watsonx.ai,** consider selecting **a lower runtime environment** to avoid high resource consumption and quickly depleting your credits. Notebook runtimes are billed based on **Capacity Unit Hours (CUH)** at a rate of **$1.02 USD per CUH.**
  Learn more about capacity unit hours and watsonx.ai Studio pricing plans.

Sample exercises:

- Prompt a foundation model with the retrieval-augmented generation pattern
- Create a LangChain AI Agent in Python using watsonx
- Create a LangChain RAG system in Python with watsonx
- Using the IBM Granite models for time series forecasting
- Build an AI stylist with IBM Granite using watsonx.ai
- Build an agentic framework with CrewAI and IBM watsonx.ai

- [Notebook: Build a LangChain agentic RAG system using the Granite model in watsonx.ai](#)

Save your Prompt Lab session

You can save your Prompt Lab editor session for later use.

1. At the top of the Prompt Lab screen, select the **Save work** dropdown button and then select the **Save as** option.
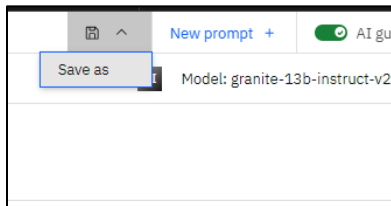
2. A **Save your work** widget will appear. Select **Prompt session** under the **Asset type** option.

3. Enter a **name** and check the **View in project after saving** option under the **Define details** section.

4. Finally, click the **Save** button. Once you save, you will see the saved work under the **Assets** tab in your sandbox project.

You can also save your work as:

- **Prompt template** to save only the current prompt without its history and selecting a **Task** suitable for your prompting.

- **Standard notebook** to continue prompting on a Jupyter Notebook environment. Prior knowledge of notebooks and Python programming language would be helpful to work with a Jupyter notebook. [Read more about notebooks](#).

Save your work on watsonx.ai

Make sure to save any work you want to retain for your records. IBM Cloud accounts will be deactivated at the end of the hackathon. Follow the steps below to save your work:

1. Go to your project's 'Overview' tab.

2. Select the 'Export or import project' drop down below the Bell icon in the top menu bar.

3. Click the 'Export project' option. This will open 'Export project to desktop' screen.

4. Select all the assets shown in your project (Work saved as Project session cannot be exported) and click 'Export' on the bottom-right of the screen.

5. The next screen will ask for confirmation that all sensitive information has been removed.

6. Click on 'Continue export'.

7. The download (zip) will be initiated and the file will be saved on your computer.

Appendix: Example use cases

You are not limited to these ideas, but here are several examples for how you could apply IBM Granite models to solve a specific issue:

- **AI Workflow Orchestrator**: Build an AI agent that understands natural language instructions and automates multi-step business workflows—like employee onboarding, approvals, or task delegation—using Granite Instruct models. This streamlines operations and reduces manual coordination.

- **Code Automation Assistant**: Create a tool powered by Granite-Code models that generates scripts or automation code from plain language prompts (e.g., "generate a Python script to clean CSV files"), enabling faster development and empowering non-coders.

- **AI-Powered Compliance Checker**: Leverage Granite Instruct models to analyze documents or communications and flag potential compliance violations or risky language, helping organizations stay aligned with regulatory standards and reduce legal exposure.

- **Smart Meeting Summarizer & Action Tracker**: Use Granite models with long-context and summarization capabilities to automatically summarize meeting transcripts and extract action items, deadlines, and responsibilities—boosting team productivity and accountability.

- **Personalized AI Learning Coach**: Develop an adaptive AI tutor using Granite models that tailors explanations, quizzes, and feedback to individual learners' styles and progress, making education more personalized and accessible.