

/ Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

WORKBOOK

SAP Generative AI Hub with AI Launchpad, Orchestration and Document Grounding

 7 Units  1 hr 45 mins  Hands-on practice

[Start Workbook](#)



Overview

Get an introduction to the AI Launchpad in SAP Generative AI Hub. Learn how to create and manage LLM prompts, and experience a practical example of Retrieval Augmented Generative (RAG) search using document grounding.

SAP Generative AI Hub with AI Launchpad, Orchestration and Document Grounding

 UNIT 1

Guided Tour

 1 Lesson  5 mins

Content

-  [Explore Demo](#)
[Guided Tour](#)

[Go to learning](#) UNIT 2

Introduction to Generative AI at SAP

 1 Lesson  5 mins

After completing this unit, you
will be able to:

Get to grips with the AI
Launchpad and its
functionalities.

Content

-  [Introduction to
Gen AI at SAP](#)

[Go to learning](#)

 UNIT 3

SAP AI Core and SAP AI Launchpad setup

 1 Lesson  10 mins

After completing this unit, you will be able to:

Setup and configure SAP Generative AI Hub using SAP AI Core and SAP AI Launchpad Boosters. SAP AI Core and SAP AI Launchpad are services which you can link to your BTP global account. SAP AI Core offers a powerful AI runtime which is natively integrated with SAP AI Launchpad. The launchpad offers an easy-to-use interface to manage AI workflow administration, processes, and tasks.

Content

-  Quick start with SAP AI Core Setup

[Go to learning](#)

 UNIT 4

Using SAP AI Launchpad

 3 Lessons

35 mins

After completing this unit, you will be able to:

SAP AI Launchpad can be used by both AI scenario producers and AI scenario consumers. AI scenario producers, such as AI operations engineers or AI engineers, are responsible for developing and productizing AI scenarios. AI scenario consumers, such as business analysts, subscribe to a service that offers an AI scenario and consume it. The generative AI hub within SAP AI Launchpad can be used to interact with generative AI models.

Content

-  [Launchpad Introduction](#)
-  [Chat](#)
-  [Prompt Editor](#)

[Go to learning](#)

 UNIT 5

Prompt Management

 1 Lesson  10 mins

After completing this unit, you will be able to:

Manage and maintain a collection of prompts

Content

 [Prompt Management](#)[Go to learning](#) UNIT 6

Prompt Engineering

 1 Lesson  20 mins

After completing this unit, you will be able to:

Design prompts in accordance with user requirements

Content

 [Prompt Engineering](#)[Go to learning](#)

 UNIT 7

SAP Generative AI Hub Orchestration Service

 2 Lessons  20 mins

After completing this unit, you will be able to:

Understand SAP Generative AI Hub Orchestration Service features and consumption methods.

Content

-  [Introduction to Orchestration](#)
-  [Orchestration with Grounding](#)

[Go to learning](#)



Orchestration with Grounding



Objectives

After completing this lesson, you will be able to:

See the effects of grounding your model

In this lesson, we are going to look at the concept of *Grounding* in terms of Generative AI. When a question is posed to a large language model, it will try to answer to the best of its knowledge, based on information it can find across the world wide web. However, it is usually much more beneficial to instruct the LLM to first consult a very specific domain before returning the results. This is known as Retrieval Augmented Generation , or RAG, whereby the results are ‘augmented’ by first grounding the LLM.

Grounding provides specialized data retrieval through vector databases, augmenting the retrieval process using specific external and context relevant data. Grounding combines generative AI capabilities with the capacity to use real-time, precise data to improve decision-making and business operations, for targeted business AI-driven solutions.

Try it out!

Note: Make sure to select the **doc-grounding** resource group as this is where the orchestration deployment with grounding is running.



Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

The screenshot shows the SAP Generative AI Hub interface. On the left, there's a sidebar with options like Model Library, Grounding Management, Chat, Prompt Editor, and a red-highlighted **Orchestration** button. The main area has two tabs: "AI API Connections (1)" and "Resource Groups (4)".

- AI API Connections:** Shows one connection named "coreai".
- Resource Groups:** Shows four groups: "rage-test" (Created On: 12 Nov 2024), "doc-grounding" (highlighted in blue, Created On: 27 Feb 2025), "grounding" (Created On: 14 Nov 2024), and "default" (Created On: 10 Apr 2024).

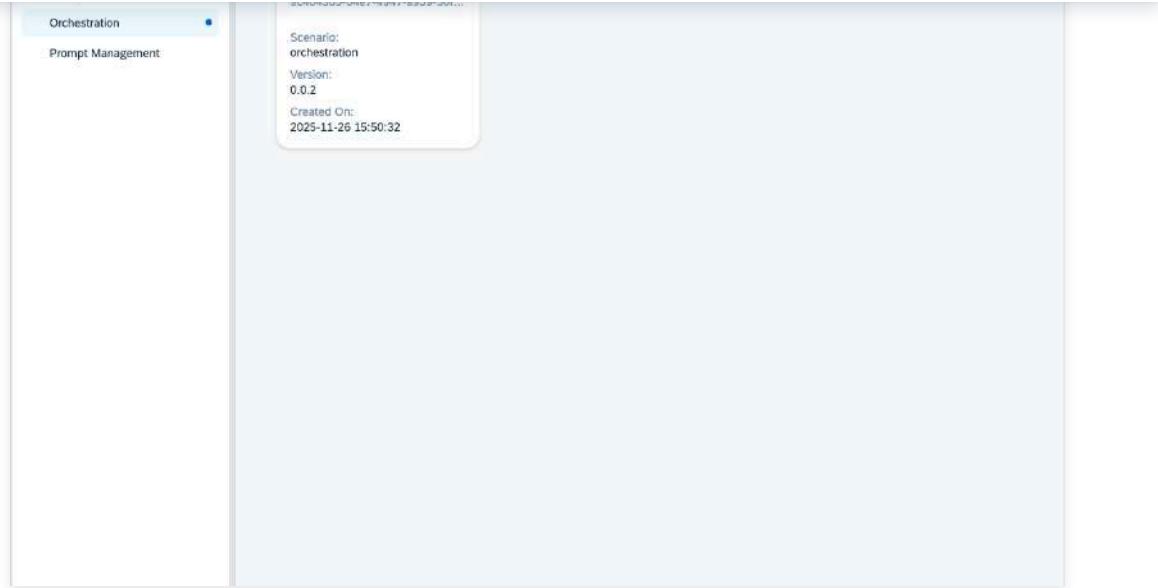
2. Expand **Generative AI Hub** and select **Orchestration** from the menu list to see the Orchestration Workflow.

This screenshot is similar to the previous one, but the "Orchestration" menu item in the sidebar is now highlighted with a red box. The main interface remains the same, showing the AI API Connections and Resource Groups sections.

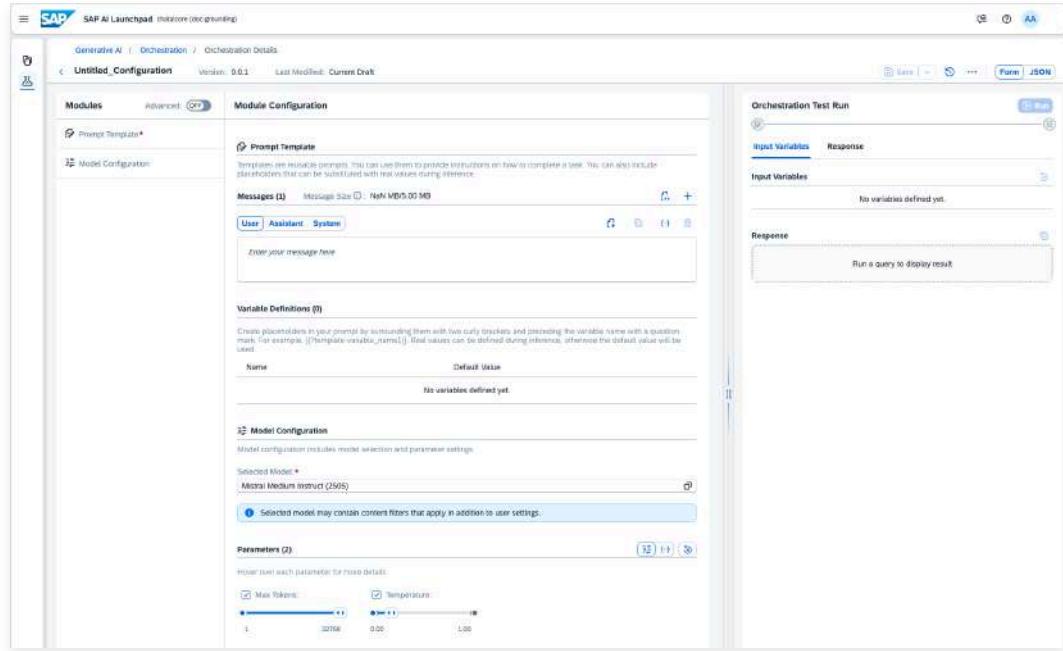
3. The *Orchestration Configurations* page will open, where any existing orchestration workflow configurations are displayed.
4. Select the **Create** button to create a new configuration.

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...



5. The *Orchestration Configuration* page will open where you can configure multiple settings for a range of different modules such as Grounding, Data Masking, Filtering and Translation.



6. Toggle the **Advanced** setting to **ON** to see the different modules available. You will notice that some of these are mandatory, while others are optional:

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

Let's first observe the effects of querying an LLM with no context and no frame of reference.

7. Ensure the *Grounding Management* feature is disabled (switch is set to Off) then go to **Prompt Template** section.

Using templates enables you to compose prompts, define placeholders, choose the LLM and configure it with required settings and parameters. It can also be enhanced with placeholders or variables to enable different prompts using the same basic configuration. Prompt Templates can be uploaded from a JSON file, the prompt library, or via the Prompt Registry using Github.

Let's upload a saved template which has been pre-configured and loaded into the prompt registry

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

Prompt Template

Templates are reusable prompts. You can use them to provide instructions on how to complete a task. You can also include placeholders that can be substituted with real values during inference.

Messages (1) Message Size 0 B/5.00 MB

User Assistant System

Enter your message here

Variable Definitions (0)

Create placeholders in your prompt by surrounding them with two curly brackets and preceding the variable name with a question mark. For example, `[?template-variable_name1]`. Real values can be defined during inference; otherwise the default value will be used.

Name	Default Value
No variables defined yet.	

The list of currently available templates will appear. It's possible to filter by the type of template (either imperative or declarative) or the search box can be used to find a specific template.

9. In the **Filter** section, select the checkbox next to **declarative** to reduce the list size. Then select the template named **prompt-registry-v1** and choose **Select** to add it to the templating module.

Select template

Prompts (0) **Templates (1)**

Filters

Managed By:

Name	Scenario	Managed By	Access Level	Created On
prompt-registry-v1 Versions: 1.1.2	product-inventory-agent	declarative	Tenant	1 Dec 2025, 09:37:52

Select **Cancel**

The template will now load up with pre-defined entries for both the **system** role and **user** role. You will also see a variable called **question** which has been defined and assigned a default value:

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

The screenshot shows the SAP Generative AI Hub interface. On the left, there's a sidebar with icons for Data Masking, Input Filtering, Output Filtering, and Output Translation. The main area has tabs for Variable Definitions, Model Configuration, and Parameters. In the Variable Definitions tab, there's a table with one row: Name 'question' and Default Value 'What types of mouse are available from the catalog?'. In the Model Configuration tab, it says 'Selected Model: Mstral Medium Instruct (2505)'. In the Parameters tab, there are two checkboxes: 'Max Tokens:' and 'Temperature:'. To the right, there are sections for Response and Trace, each with a button to 'Run a query to display result'.

10. Scroll down to the **Model Configuration** section to choose a model from the available versions and define its parameters. A model has already been configured for this step, but you can choose from a large selection of other available models in here.

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

You are a helpful AI assistant for SAP AI Core. Answer the question by providing a relevant answer only from Reports that fit to the request.

User Assistant System

Take the user's Request: {{ ?question }} from here

Variable Definitions (1)

Create placeholders in your prompt by surrounding them with two curly brackets and preceding the variable name with a question mark. For example, {{?template-variable_name1}}. Real values can be defined during inference, otherwise the default value will be used.

Name	Default Value
question	What types of mouse are available from the catalog?

Model Configuration

Model configuration includes model selection and parameter settings.

Selected Model:*

Mistral Medium Instruct (2505)

Selected model may contain content filters that apply in addition to user settings.

Parameters (2)

Hover over each parameter for more details

Max Tokens:



Temperature:



1.00

11. Open up the list of available models and choose **SAP ABAP 1** for this exercise. Selecting the model will automatically add it to the *Model Configuration* section where further parameters can be tweaked.

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

The screenshot shows a grid of 12 LLM models listed in three rows of four. Each model card includes its name, provider, version, and a set of configuration icons.

Model Name	Provider	Version	Actions
(gemini-2.5-flash)		Version: 001	
(gemini-2.5-flash-lite)		Version: 001	
(mistralai--mistral-large-instruct)		Version: 2407	
Mistral Medium Instruct	(mistralai--mistral-medium-instruct)	Version: 2505	
Mistral Small Instruct	(mistralai--mistral-small-instruct)	Version: 2503	
aws Nova Lite	(amazon--nova-lite)	Version: 1	
aws Nova Micro	(amazon--nova-micro)	Version: 1	
aws Nova Pro	(amazon--nova-pro)	Version: 1	
o3	(o3)	Version: 2025-04-16	
o3 Mini	(o3-mini)	Version: 2025-01-31	
o4 Mini	(o4-mini)	Version: 2025-04-16	
Perplexity Sonar	(sonar)	Version: perplexity-us	
Perplexity Sonar Pro	(sonar-pro)	Version: perplexity-us	
SAP ABAP 1	(sap-abap-1)	Version: 1	

Once happy with the model selection and configuration, it is time to test out the prompt.

On the right hand side, you will see the **Orchestration Test Run** section, which has already been populated with the input variable from the template.

11. Select **Run** to send the query to your chosen LLM and observe the response.

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

The screenshot shows the SAP Generative AI Hub interface. At the top, there are tabs: Input variables (selected), Message History (1), Response, and Trace. The Input Variables section contains a red-bordered input field labeled "question:" with the text "What types of mouse are available from the catalog?". The Message History (1) section has tabs for User, Assistant (selected), and System, with a placeholder "Enter your prompt message here...". The Response section contains a dashed box with the text "Run a query to display result". The Trace section also contains a dashed box with the same text.

As this example uses a very targeted query with no context (looking for items in a specific catalog), the response from the model will either be an admission of failure to find the items, or it will make up an answer with generic options. This will depend on the model and version chosen, along with the parameters defined.

12. Save your current configuration by selecting the **Save** button at the top of the screen.

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

The screenshot shows the SAP Generative AI Hub interface. At the top, there's a header bar with the SAP logo and navigation links. Below it, a breadcrumb trail indicates the current location: Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

The main content area is divided into several sections:

- Input Variables:** A section for defining variables, currently empty.
- Message History (1):** A table showing a single message entry.

User	Assistant	System

A text input field below the table says "Enter your prompt message here..."
- Response:** A large text area containing a response to a user query.

I don't have the catalog or Reports you're referring to. Please either:
 - Attach or upload the catalog/report file (PDF, CSV, XLSX, or plain text), or
 - Paste the catalog text or a link to the report, or
 - Tell me which specific Report name (from your system) I should read.Once you provide the report, I will extract and list the mouse types available (and can also give counts, SKUs, or brief specs on each if you want).
- Trace:** A section for viewing the execution trace of the orchestration.

At the bottom of the interface, there are some token statistics: Prompt Tokens: 55 and Response Tokens: 560.

13. Give your configuration the name **GE275967_Orch**, select **orchestration** as the Scenario Name and then select **Save**.

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

The screenshot shows a configuration dialog box. At the top, it says "Orchestration Configuration Name: *". Below that is a text input field containing "AC12345U01_Orch". Underneath is another field labeled "Scenario Name: *". This field has a dropdown menu open, showing the option "orchestration". At the bottom right of the dialog are two buttons: a blue "Save" button with a white border and a white "Cancel" button with a light gray border. A red rectangular box highlights the "Save" button.

In the next section, we can offer the model more context for the query by providing it with a database of information from which to search before providing an answer.

Grounding

Let's now 'ground' the LLM on a repository of data before asking the same question as before.

For this scenario, we have already created a set of embeddings from a product catalog which have been stored as vectors in a HANA Cloud Database.

1. Select the **Grounding Management** menu option within the *Generative AI Hub* section to view the available data repositories which can be used in this resource group (*doc-grounding*).

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

The screenshot shows the SAP Generative AI Hub interface. On the left, there's a sidebar with 'Prompt Management' and 'Optimization' sections, both currently set to 'Disabled'. Under 'Prompt Management', there are six options: 'Grounding Management' (OFF), 'Input Translation' (OFF), 'Data Masking' (OFF), 'Input Filtering' (OFF), 'Output Filtering' (OFF), and 'Output Translation' (OFF). In the center, there's a 'Messages (2)' section with a message size of 191 B / 5.00 MB. The first message is from 'User' and says: 'You are a helpful AI assistant for SAP AI Core. Answer the question by providing a relevant answer only from Reports that fit to the request.' The second message is also from 'User' and says: 'Take the user's Request: {{ ?question }} from here'. Below these messages is a 'Variable Definitions (1)' section with one entry: 'question' with a default value of 'What types of mouse are available from the catalog?'. At the bottom right, there's a vertical scroll bar.

This is a collection of vectorized representations of various documents, csv files, images etc. which have been stored in MS Sharepoint, S3 buckets or directly in the vector store of SAP HANA Cloud itself.

Note: The repositories that you see may differ from those in the image below as this can depend on which tenant you are in and how many repositories have been pre-configured.

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

	Prompt Editor Orchestration Prompt Management	pipeline-158f141d-42e5-454a-adf1-c374b3f0f504-collection 2774102a-0a0a-4498-a636-91... Type: MSSharePoint	pipeline-7e0bc74f-cfe0-449c-8f18-85b46dc9e40b-collection 376833ec-20d1-4828-bd22-23... Type: MSSharePoint	genaihub-vector be9bb510-f47e-46b0-b068-ae0... Type: Vector DB
		pipeline-c9138100-2bc1-4bc6-8731-72137bb8a583-collection cc4c2ce6-e3cb-4b61-a920-b00... Type: MSSharePoint	pipeline-b02419ff-ff04-44e7-86d3-4d81bbc9dcc-collection 45264529-037-4910-90ef-ced... Type: MSSharePoint	pipeline-725fd0ac-d1d5-49df-8b38-af3cf3c75d66-collection 2a279c72-0434-4ed5-98e8-2b... Type: S3
		pipeline-aeac815a-3079-465b-bb66-63709b03c149-collection e004566f-5abb-44b6-8f4a-21f7... Type: S3	pipeline-3ed0ce37-4556-4675-82f7-b45be907e959-collection 5d624ce1-b5f3-4af5-b8d6-4e1... Type: MSSharePoint	pipeline-f2d9e4a2-7c25-476c-abd8-4bad5271bcfe-collection 92e85424-2317-4315-be7c-79... Type: MSSharePoint

2. Select any of the repositories to view the data chunks. The repository we use for this workshop is called **pipeline-86a82ae8-e3a6-4f68-97b1-63568965f174-collection**. Select to open the data and observe the format.

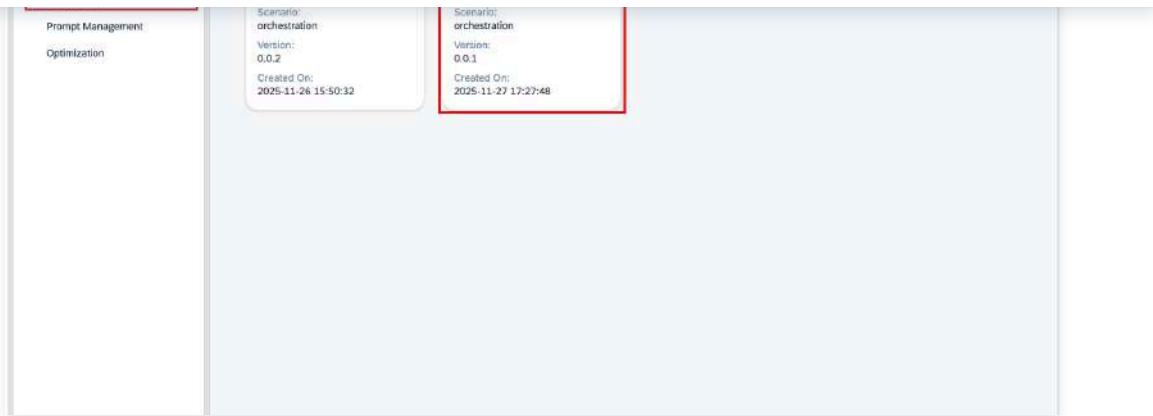
The screenshot shows the SAP Generative AI Hub interface. On the left, there's a sidebar with options like Workspaces, Generative AI Hub, Model Library, Grounding Management, Chat, Prompt Editor, Orchestration, and Prompt Management. The main area is titled "Generative AI / Grounding Management / Data Repository Details" and shows a collection named "pipeline-c9138100-2bc1-4bc6-8731-72137bb8a583-collection". It displays "Resources (2)" and "Documents" (with two entries: 36a1d125-c488-53de-8fc5... and 142e2bb-9856-5a19-807d...). Below this is a "Document Details" section showing a single item with ID: cc4c2ce6-e3cb-4b61-a920-b00... and Type: MSSharePoint. The bottom part of the screen shows a list of "Chunks (6)" and "Metadata (12)". The "Chunks" section lists several items, including:

- 36a1d125-c488-53de-8fc5-71b2ef8cb2d
- 3fed10c5-c4b2-53de-8fc5-71b2ef8cb2d
- 7.41315782770495 51.51424727981876 8 Dortmund P_0117 Ergonomic Keyboard IT Accessories 10900 HD webcam with privacy shutter 239.38 Each 5191 SpeedStorage 1.10 EU...
- 5125 SpeedStorage 9.49 EURO Germany Südwald, 44137 Dortmund, Germany 7.407332872246005 51.5106189930273 9 Dortmund 709 P_0114 Wireless Mouse IT Accessories Blue...

3. Download the following [JSON template file](#), and save it locally on your device. Feel free to inspect the file and observe the format and layout.
4. Return to the orchestration workflow by selecting **Orchestration** from the menu and then choose your previously created configuration called **GE275967_Orch**.

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...



5. In the orchestration configuration screen, select the three dots at the top right and then choose **Upload** which will lead to a pop-up window asking for a file to be uploaded.

6. Choose the JSON file which you have saved in a previous step, and select **Open** to load it into the workflow.

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...



7. Observe that the *Grounding Management* module has now been activated and is enabled. There are also two new additions in the configuration section: An **Output Variable** and a **Data Repository**.

Name	Default Value	Max Chunks
pipeline_6fa	Max Document	1

Note: The Data Repository is the vectorized representation of an internal catalog of IT equipment and this forms the basis on which the LLM should formulate its response.

The addition of an output variable tells the LLM to use the information retrieved from the data repository and store it in an output variable which can be used in a template.

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

Create placeholders in your prompt by surrounding them with two curly brackets and preceding the variable name with a question mark. For example, {{?template-variable_name1}}. Real values can be defined during inference, otherwise the default value will be used.

Name	Default Value
grounding_output_variable	
question	What types of mouse are available from the catalog?

Model Configuration

Model configuration includes model selection and parameter settings.

Selected Model: *

GPT-5-Mini (2025-08-07)



i Selected model may contain content filters that apply in addition to user settings.

Parameters (2)



Hover over each parameter for more details

Max Completion Tokens:



Temperature:



This is because the JSON configuration file has a section in which you can define the model and its parameters. In fact, there are many different options available to define in this configuration file, and it is an effective way of storing different orchestration workflows.

9. Select the JSON icon to view the JSON representation of the orchestration workflow. In particular, observe the addition of the data repository used for grounding:

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

The screenshot shows a JSON configuration file on the left and a response window on the right.

```

1: {
  "question": "What types of mouse are available from the catalog?",
  "output": "grounding_output_variable"
}
2: {
  "prompt_template": {
    "template": {
      "role": "system",
      "content": [
        {
          "type": "text",
          "text": "You are a helpful AI assistant for SAP AI Core. Answer the question by providing a relevant answer only from Reports that fit to the request. Reports: {{? grounding_output_variable}}"
        }
      ]
    },
    "role": "user",
    "content": [
      {
        "type": "text",
        "text": "Take the user's Request: {{question}} from here"
      }
    ]
  },
  "defaults": {
    "question": "What types of mouse are available from the catalog?"
  }
},
3: {
  "model": {
    "name": "gpt-3-miri"
  }
}

```

The response window shows the message: "User Assistant | System" and "Enter your prompt message here...". Below it, the "Response" section is empty.

- When ready, select the **Run** button again to test the query on a model which is now grounded on the IT equipment data repository and observe the results.

The screenshot shows the AC272139U01 Orchestrator interface with two main sections: Configuration and Test Run.

Configuration (Left):

- Modules:** Advanced (On)
- Activated:**
 - Grounding Management (On)
 - Prompt Template (On)
 - Model Configuration (On)
- Disabled:**
 - Input Translation (Off)
 - Data Masking (Off)
 - Input Filtering (Off)
 - Output Filtering (Off)
 - Output Translation (Off)
- Variable Definitions (2):**

Name	Default Value
grounding_output_variable	
question	What types of mouse are available from the catalog?
- Model Configuration:**
 - Selected Model: GPT-3-Miri (2025-08-07)
 - A note: Selected model may contain content filters that apply in addition to user settings.
- Parameters (2):**
 - (None)

Test Run (Right):

- Orchestration Test Run:** Run button (On)
- Input Variables:** question: What types of mouse are available from the catalog?
- Message History (1):** User: Enter your prompt message here...
- Response:** Run a query to display result
- Trace:** Run a query to display result

- This time you should get a more targeted response suggesting the model was much more aware of the context behind the query:

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

Response

All catalog entries labeled as mouse are listed as "Wireless Mouse." Details for each entry:

- P_0111 — Wireless Mouse
 - Supplier: SpeedStorage (Germany, Frankfurt)
 - Unit price: 34.75 EUR
 - Stock: 104
 - Lead time: 13 days
 - Description field: "Bluetooth ergonomic keyboard with customizable keys"
- P_0138 — Wireless Mouse
 - Supplier: MouseWare (Germany, Cologne)
 - Unit price: 147.25 EUR
 - Stock: 498
 - Lead time: 2 days
 - Description field: "USB-C docking station with multiple ports"
- P_0052 — Wireless Mouse
 - Supplier: StandSolutions (Germany, Dortmund)
 - Unit price: 295.93 EUR
 - Stock: 735
 - Lead time: 13 days
 - Description field: "Bluetooth ergonomic keyboard with customizable keys"

Note: All items are named "Wireless Mouse" in the catalog, but the description fields contain

Prompt Tokens: 364 Response Tokens: 622

Trace

This simplified example showcases how the grounding of a foundation model before running queries enhances its performance and reliability by aligning it with specific context and domain knowledge. This process involves customizing and fine-tuning the model using relevant data, which improves its understanding of nuanced language and concepts pertinent to specific applications.

Benefits include:

- Increased accuracy
- More relevant query results
- Reduced biases
- Better handling of specialized vocabularies

Ultimately, grounding ensures that the model's outputs are more precise and contextually appropriate, leading to higher overall effectiveness in various tasks like information retrieval, decision support, and predictive analytics.

Basic Trials

[Home](#) / [Workbooks](#) / SAP Generative AI Hub with AI Launchpad, Orchestration and Document ...

SAP Business Technology Platform

Basic Trials



[Home](#) / [Workbooks](#) / SAP Generative AI Hub with AI Launchpad, Orchestration and Document G...



Introduction to Orchestration



Objectives

After completing this lesson, you will be able to:

Understand SAP Gen AI Hub orchestration service

In the context of AI, orchestration involves coordinating various AI models, data sources, and computational resources to achieve a specific goal.

Key Functions:

- **Templating**
- **Grounding**
- **Translation**
- **Content filtering**
- **Data Masking**
- **Orchestration Workflow**
- **Configuration and Execution**
- **Harmonized API**

Orchestration Workflow

In a basic orchestration setup, you can combine different modules into a pipeline executed with a single API call. Each module's response serves as the input for the next module.



Basic Trials

[Home](#) / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document G...

Grounding

Grounding integrates external, contextually relevant, domain-specific, or real-time data into AI processes. This data supplements the natural language processing capabilities of pre-trained models, which are trained on general material.

Grounding is a service designed to handle data-related tasks, for example grounding and retrieval, using vector databases.

The Pipeline API is proxied through the SAP AI Core generative AI hub, and incorporates vector stores, such as the managed SAP HANA Cloud database.

Grounding provides specialized data retrieval through vector databases, grounding the retrieval process using your own external and context relevant data. Grounding combines generative AI capabilities with the capacity to use real-time, precise data to improve decision-making and business operations, for specific business AI driven solutions.

Templating

The templating module is mandatory. It enables you to compose prompts and define placeholders. It then generates the final query that is sent to the model configuration module. Any placeholders that you define in your prompt can be filled at runtime. For example, the following template places the input text into the `{{ ?input }}` placeholder, for example:

```
"templating_module_config": {  
    "template": [  
        {  
            "role": "user",  
            "content": "{{ ?input }}"  
        }  
    ]  
}
```

The `{{ ?input }}` placeholder is filled using the contents of the `input_params` input parameter:

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document G...

You can also set default values for the placeholders. For example, to request either 5 paraphrases or a user-specific number of paraphrases for a given phrase, you can use the following configuration:

```
"templating_module_config": {
    "template": [
        {
            "role": "user",
            "content": "Create {{ ?number }} paraphrases"
        }
    ],
    "defaults": {
        "number": 5
    },
    //... other configuration
    "input_params": {
        "number": "3",
        "phrase": "Please respond as soon as possible, th
    }
}
```

As you will see in the next lesson, it's possible to upload a pre-defined template using a JSON config file, or via the [Prompt Registry](#).

Data Masking

The data masking module is optional. It anonymizes or pseudonymizes personally identifiable information from the input.

- Anonymized data can not be unmasked as information about the original data is not retained. Personally identifiable information is replaced with a **MASKED_ENTITY** placeholder.
- Pseudonymized data can be unmasked in the response. personally identifiable information is replaced with a **MASKED_ENTITY_ID** placeholder.

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document G...

The masking service can mask personally identifiable information in the prompt. However, because it is using automated detection mechanisms, there is no guarantee that all personally identifiable information will be found and masked.

Anonymization, replaces personally identifiable information in an irreversible manner. This results in a loss of context, which may limit the capability of the LLM to process the input.

For example, if asked to write a story about Ben and Anna, with anonymization of profile-person, the LLM would be asked to write a story about **MASKED_PERSON** and **MASKED_PERSON** and could no longer distinguish between the two.

Input Filtering

The content filtering module is optional. It lets you filter the input and output based on content safety criteria.

The module supports the Azure Content Safety classification service. This service recognizes four distinct content categories: Hate, Violence, Sexual, and SelfHarm. For more information, see Harm categories in Azure AI Content SafetyInformation published on non-SAP site. Text can have more than one label (for example, a text sample can be classified as both Hate and Violence). The returned content categories include a severity level rating of 0, 2, 4, or 6. The value increases with the severity of the content.

If no content filter is configured in the orchestration configuration, the input is passed to the model configuration without filtering. Similarly, the response is returned without filtering. If no severity levels are set in the configuration (i.e. the config key is not given), the default level of 2 (low) is used for all categories.

Model Configuration

Basic Trials

Home / Workbooks / SAP Generative AI Hub with AI Launchpad, Orchestration and Document G...

- Model Parameters: The parameters to be applied to the model.
- Model Version: The version of the model to be used, which can either be latest or a specific version number.

Output Filtering

The content filtering module is optional. It lets you filter the input and output based on content safety criteria.

The module supports the Azure Content Safety classification service. This service recognizes four distinct content categories:

- Hate
- Violence
- Sexual
- SelfHarm.

If no content filter is configured in the orchestration configuration, the input is passed to the model configuration without filtering. Similarly, the response is returned without filtering. If no severity levels are set in the configuration (i.e. the config key is not given), the default level of **2 (low)** is used for all categories.

Harmonized API

The harmonized API lets you use different foundation models without the need to change the client code. It does so by taking the OpenAI API as the standard and mapping other model APIs to it. This includes standardizing message formats, model parameters, and response formats. The harmonized API is integrated into the templating module, the model configuration, and the orchestration response.

In the OpenAI format, a prompt and its response can contain a list of messages with a role and content. The role can be system, user, or assistant and the content is the text of the message. The response can contain a list of choices with **index**, **messages**, and a **finish_reason**.

For more information see [Harmonized API](#)