



data
iku

KNOWLEDGE

[Home](#) » [Machine Learning & Analytics](#) » [Generative AI and Large Language Models \(LLMs\)](#)

Concept | Prompt Studios and Prompt recipe

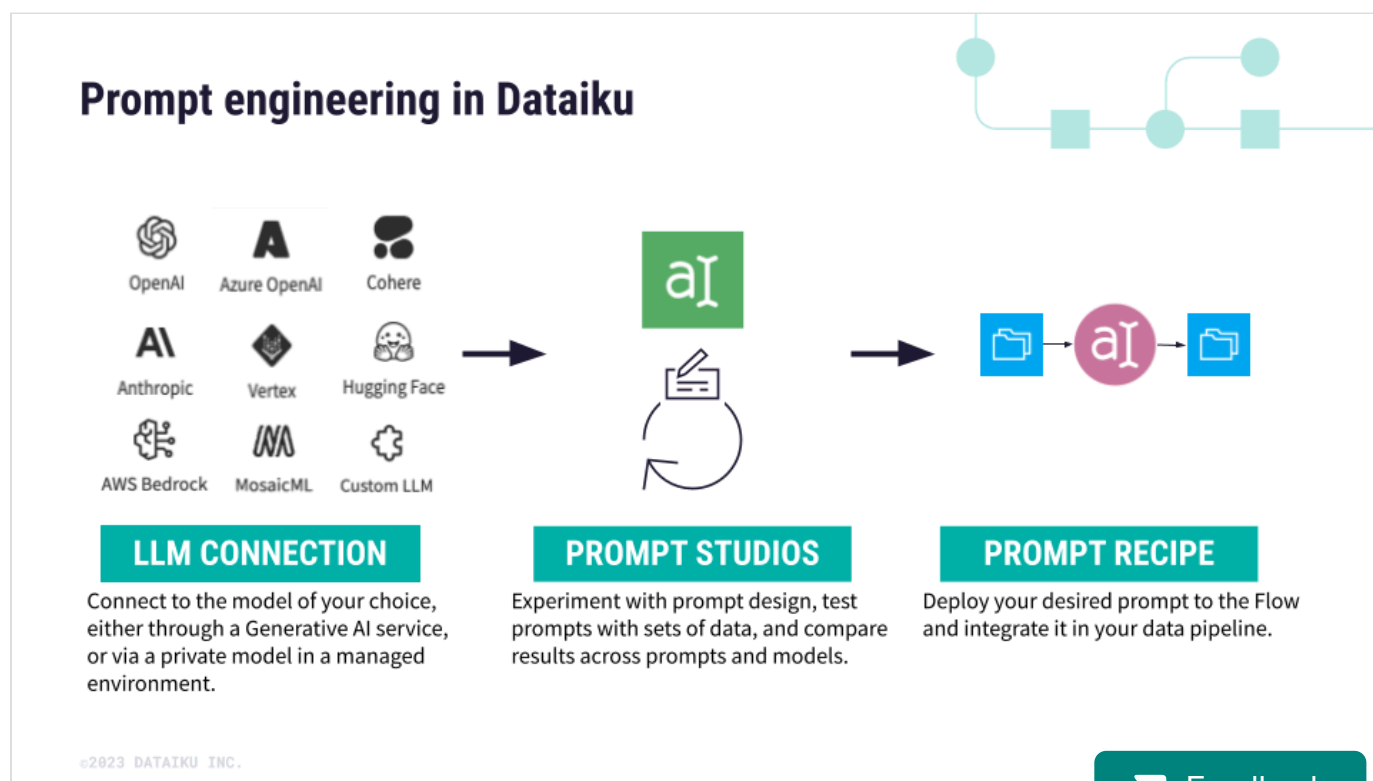
Prompts are the input to a Generative AI model such as a Large Language Model (LLM). You can modify and craft prompts to achieve a desired outcome — a process known as **prompt engineering**.

Prompts guide the model's behavior and output, but they don't modify the underlying model.

Prompt engineering in Dataiku

In Dataiku, **Prompt Studios** and the **Prompt recipe** work together to test and iterate on LLM prompts, then operationalize them in your data pipeline.

Prompt Studios and the Prompt recipe allow you to securely connect to LLMs either through a Generative AI service or a private model, including open-source models. You can first experiment with different prompt designs in Prompt Studios, before deploying a desired prompt in your data pipeline with the Prompt recipe.



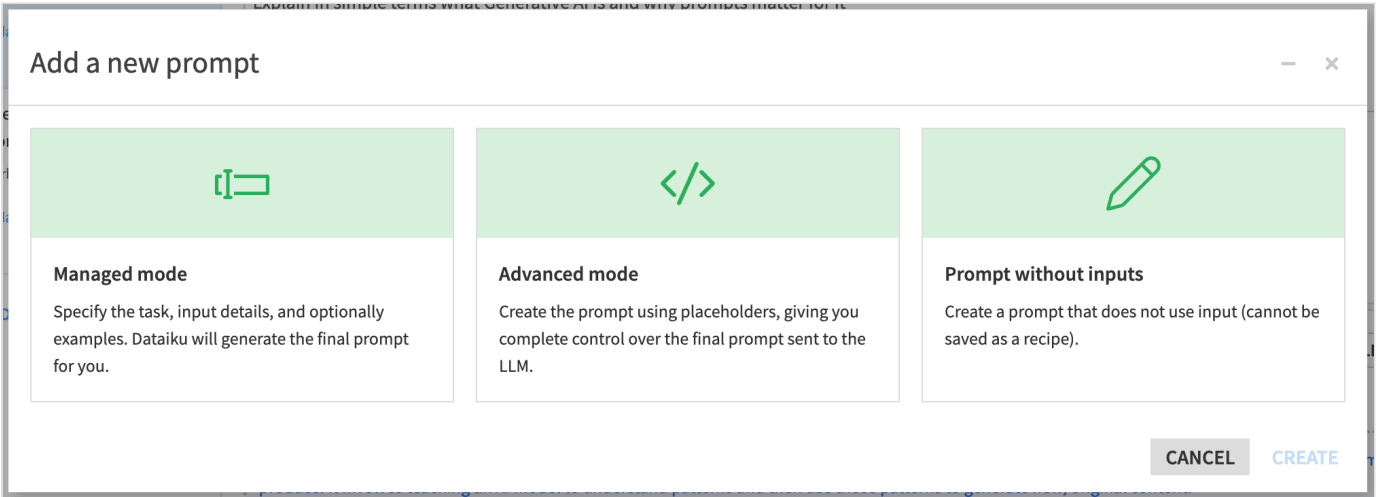
Feedback

Prompt Studios

Prompt Studios is the playground where you can iterate and test different prompts and models to engineer the optimal prompt for a specific use case. You can find Prompt Studios in the top navigation bar under the Visual Analyses menu.

Within a Prompt Studio, you can create multiple prompts and iterations, and even compare them. For each new prompt, you can choose from three different modes:

Mode	Description
Managed mode	You specify the task using plain language, optionally include other details, and Dataiku generates the final prompt for you. You can use templates for common tasks.
Advanced mode	You create the prompt and examples using placeholders, giving you more control over the final prompt sent to the LLM.
Prompt without inputs	You create a one-off prompt that queries the LLM and is designed for quick experimenting. These prompts can't be converted into a Prompt recipe and their results cannot be compared in Prompt Studio.



Designing prompts

The **Prompt design** steps will differ depending on the mode you're using. In general, this process includes the elements described below.

Element	Description
Prompt	Includes the main instructions to the LLM.
Inputs	Lists the variables inserted into the prompt. You can map the inputs to columns from a dataset and use values from the dataset as test cases to check the prompt's output. You can also choose to manually write test cases.
Examples	Provides sample input-output pairs to guide the model's expected results.
Test cases	Evaluates prompt performance using either dataset rows or manually written examples.

You can quickly iterate on your prompt design, including updating the task or adding new inputs, examples, and test cases, before deploying your prompt.

Each time you run inference on your prompt and test cases, the model will also return an estimated cost for 1,000 records, helping you manage costs associated with LLM usage.

The screenshot displays the Prompt Studio interface for a project titled "Financial headlines". The interface is divided into several sections:

- Left Sidebar:** Contains a "COMPARE" button, a prompt preview "Give the topic of each article based on the headline and short description." using GPT-4o with an estimated cost of \$0.49/1k records, and an "ADD PROMPT" button.
- Prompt design section:**
 - LLM:** Set to GPT-4o (connection: open-ai-carolin).
 - Prompt:** "Give the topic of each article based on the headline and short description."
 - Inputs from:** A dropdown menu showing "reuters_headlines" (DATASET - View).
 - Input Mapping:** Two inputs are defined: "headline" mapped to the "Headlines" column and "description" mapped to the "Description" column.
 - Buttons:** "ADD INPUT" and "ADD IMAGE INPUT".
- Examples section:** Shows a table with input-output pairs.

Input: headline	Input: description	Output
U.S. economy faces significant risks	The U.S. economy is forecast to shrink	economy, pandemic
- Test cases section:**
 - Summary: 8 records, 8 OK (no validation rules), Estimated cost: \$0.49/1k records.
 - Buttons:** "CLEAR" and "RUN PROMPT".
 - Table:**

Headlines	Description	Result from last run
TikTok considers London and other locations for headquarters	TikTok has been in discussions with the UK government over the past few months to locate its headquarters in London, a source familiar with the matter said, as part of a strategy to distance itself from its Chinese ownership.	business, technology, corporate relocation
Disney cuts ad spending on Facebook amid growing boycott: WSJ	Walt Disney has become the latest company to slash its advertising spending on Facebook Inc as the social media giant faces an ad boycott over its handling of hate speech and controversial content, the	social media, advertising boycott

Using images in prompts

You can also process images within Prompt Studios for use cases where visual input is useful.

To do this, you will need:

- Dataiku version 13.2 or above.
- A connection to an LLM that supports images as inputs.
- Images stored in a managed folder, which you can connect to within Prompt Studios. (See [How to | Prepare images for use in a model](#) for instructions.)

The screenshot shows the Dataiku Prompt Studio interface for a task named 'classify dog breeds'. The interface is divided into several sections:

- Left Panel:** Contains a 'COMPARE' button, a prompt description 'Give the breed of each dog shown in the images.', and a selection for 'GPT-4o' with a cost of '\$0.82/1k records'. There is also an '+ ADD' button.
- Center Panel:** Labeled 'Prompt design', it shows the LLM 'GPT-4o (connection: open-ai-caroli)' and a 'Prompt' section with the instruction 'Define the task that the prompt must fulfill'. Below this is a large image of a dog in a room with framed pictures.
- Right Panel (Inputs):** A red box highlights the 'Inputs' section. It shows 'Inputs from' set to 'dog_breed_images_files' (a dataset) and 'Image folder' set to 'dog_breed_images' (a managed folder). Below this, there are fields for 'Description' (set to 'image') and 'Column' (set to 'path'). There are also '+ ADD INPUT' and '+ ADD IMAGE INPUT' buttons.
- Bottom Panel:** Shows the 'Result from last run' with four entries: 'The dogs in the image appear to be a Boston Terrier (black and white) and a Chihuahua (tan).', 'This dog appears to be a Beagle.', 'The dog in the image appears to be a Golden Retriever.', and 'This dog appears to be a Border Collie.' There are also 'CLEAR' and 'RUN PROMPT' buttons.

Settings

You can change several hyperparameters or set validation rules for the prompt.

Tab	Description
LLM settings	Changes hyperparameters such as temperature or the maximum output tokens, that control the behavior of the LLM.
Validation & formatting	Sets rules such as validating the expected output format as a JSON object or forbidding certain words from the output. You will receive a warning if the output is not validated.

Prompt settings

LLM SETTINGS

VALIDATION & FORMATTING

Tune the behaviour of the LLM used via the advanced settings below. Default values are used for fields left blank.

Temperature

Parameter governing the randomness in the responses, 0 for more predictable, 0.5 to 1.0 for balanced creativity.

Top P

Sample from the top tokens whose probabilities add up to p.

Max output tokens

Maximum tokens to output per row (in English, 1 word ~ 1.37 tokens)

Frequency penalty

Penalty applied on the next token proportional to how many times that token already appeared in the response and prompt.

Presence penalty

Penalty to apply on repeated tokens, no matter how many times the token already appeared in the response and prompt.

JSON mode

☐

Request the model to generate a valid JSON response, for models that support it.

CLOSE

SAVE

Comparing prompts

Prompt Studios track your prompts in the left sidebar. You can navigate between prompts, view each prompt's history, restore previous versions, and duplicate prompts.

When using managed or advanced modes, you can also compare the output of multiple prompts by selecting them in the sidebar and clicking **Compare**. Prompt Studios will build a table comparing the outputs and costs of the selected prompts.

Prompt Studios
←

COMPARISON	Comparing			
<input checked="" type="checkbox"/> COMPARE <input checked="" type="checkbox"/> Give the topic of each article based on the headline and short description. <div style="display: flex; align-items: center;"> <div style="margin-right: 10px;"></div> <div>GPT 3.5 Chat (Turbo / ChatGPT)</div> </div> <div style="display: flex; justify-content: space-between; margin-top: 10px;"> 8 OK (no validation rules) - \$0.247/1K records ☆ </div>	<div style="text-align: right; border: 1px solid blue; padding: 2px 5px; color: blue;">CLOSE COMPARISON</div>		<p>Give the topic of each article based on the headline and short description.</p> <div style="display: flex; align-items: center;"> openai:maxime_openAi:gpt-3.5-turbo </div> <div style="display: flex; justify-content: space-between; margin-top: 10px;"> 8 OK (no validation rules) - \$0.247/1K records </div>	<p>Determine whether each topic of the following list of topics is covered in the financial news article provided. List of topics: fed and central banks, company and product news, corporate debt and ear...</p> <div style="display: flex; align-items: center;"> openai:maxime_openAi:gpt-3.5-turbo </div> <div style="display: flex; justify-content: space-between; margin-top: 10px;"> 8 OK (no validation rules) - \$0.324/1K records </div>
<input checked="" type="checkbox"/> Determine whether each topic of the following list of topics is covered in the financial news article provided. List of topics: fed and central banks, company and product news, corporate debt and ear... <div style="display: flex; align-items: center;"> <div style="margin-right: 10px;"></div> <div>GPT 3.5 Chat (Turbo / ChatGPT)</div> </div> <div style="display: flex; justify-content: space-between; margin-top: 10px;"> 8 OK (no validation rules) - \$0.323/1K records ☆ </div>	<div style="display: flex;"> <div style="flex: 1;"> Headlines TikTok considers London and other locations for headquarters </div> <div style="flex: 1;"> Description TikTok has been in discussions with the UK government over the past few months to locate its headquarters in London, a source familiar with the matter said, as part of a strategy to distance itself from its Chinese ownership. </div> </div>		<p>Corporate headquarters and TikTok</p>	<p>Topics covered in the article:</p> <ul style="list-style-type: none"> - Company and product news (TikTok considering London and other locations for headquarters) - Politics (discussions with the UK government) - Legal and regulation (strategy to distance itself from its Chinese ownership) <p>Topics not covered in the article:</p> <ul style="list-style-type: none"> - Fed and central banks - Corporate debt and earnings - Energy and oil - Currencies - Gold and metals - IPO - M&A and investments - Markets - Stock movement
<div style="text-align: center; color: blue; font-weight: bold;">+ ADD PROMPT</div>	<div style="display: flex;"> <div style="flex: 1;"> Disney cuts ad spending on Facebook amid growing boycott: WSJ </div> <div style="flex: 1;"> Walt Disney has become the latest company to slash its advertising spending on Facebook Inc as the social media giant faces an ad boycott over its handling of hate speech </div> </div>		<p>Advertising boycott on Facebook</p>	<p>Topics covered in the article: company and product news, corporate debt and earnings, legal and regulation, M&A and investments, politics</p>

Tip

Prompt Studios also tracks each run of a prompt as you iterate. To view the version history of a prompt and revert back to any version, click on the clock icon in the top right of the Prompt design screen.

Prompt recipe

The Prompt recipe puts your prompt into action on a dataset. It generates an output dataset and appears in the Flow.

You can create a new Prompt recipe directly from Prompt Studios by saving a prompt template that is mapped to a dataset as a recipe. This allows you to experiment with your prompts before operationalizing them with the recipe.

You can also create a new Prompt recipe directly from the Flow or from the Actions panel of a dataset. With this method, you can write a Structured prompt or a Text prompt directly in the recipe, with the same settings as a prompt template in a Prompt Studio.

The screenshot shows the 'Prompt and LLM settings' configuration page for a recipe named 'compute_headline_summaries'. The interface includes a top navigation bar with tabs for 'Settings', 'Input / Output', 'Advanced', and 'History'. A 'SAVED!' status indicator is visible in the top right corner. The main configuration area is divided into several sections:

- LLM:** A dropdown menu is set to 'GPT 3.5 Chat (Turbo / ChatGPT) - I'.
- Prompt:** A text area contains the prompt: 'Give the topic of each article based on the headline and short description.' A 'SHOW RAW PROMPT' button is located to the right.
- Inputs:** A table maps input columns to the prompt. The first row maps 'headline' to 'Headlines', and the second row maps 'description' to 'Description'. There are trash icons next to each row and a '+ ADD INPUT' button below.
- Examples:** A table provides input-output examples. The first example has 'Input: headline' with the value 'Lured by rising SUV sales, automake' and 'Output' with the value 'Automotive markets'. The second example has 'Input: description' with the value 'Demand for sport utility vehicles in t'. There is an '+ ADD EXAMPLE' button below.
- LLM completion settings:** Includes fields for 'Temperature' (with a description: 'Parameter governing the randomness in the responses, from less (0) to more (2) creative.'), 'Top P' (with a description: 'Sample from the top tokens whose probabilities add up to p.'), and 'Max output tokens' (with a description: 'Maximum tokens to output per row (in English, 1 word ~ 1.37 tokens)').
- Result validation settings:** This section is partially visible at the bottom.

A green 'RUN' button is located at the bottom left of the configuration area. On the right side of the interface, there is a vertical sidebar with icons for adding new elements, information, chat, and a refresh/clear action.

What's next?

Continue learning about prompt engineering with LLMs by working through the [Tutorial | Prompt engineering with LLMs](#) article.