

✓ 100 XP

Use prompts to get completions from models

3 minutes

Once the model is deployed, you can test how it completes prompts. A prompt is the text portion of a request that is sent to the deployed model's completions endpoint. Responses are referred to as *completions*, which can come in form of text, code, or other formats.

Prompt types

Prompts can be grouped into types of requests based on task.

Expand table

Task type	Prompt example	Completion example
Classifying content	Tweet: I enjoyed the trip. Sentiment:	Positive
Generating new content	List ways of traveling	1. Bike 2. Car ...
Holding a conversation	A friendly AI assistant	See examples
Transformation (translation and symbol conversion)	English: Hello French:	bonjour
Summarizing content	Provide a summary of the content {text}	The content shares methods of machine learning.

Task type	Prompt example	Completion example
Picking up where you left off	One way to grow tomatoes	is to plant seeds.
Giving factual responses	How many moons does Earth have?	One

Completion quality

Several factors affect the quality of completions you'll get from a generative AI solution.

- The way a prompt is engineered. Learn more about [prompt engineering here](#).
- The model parameters (covered next)
- The data the model is trained on, which can be adapted through [model fine-tuning with customization](#)

You have more control over the completions returned by training a custom model than through prompt engineering and parameter adjustment.

Making calls

You can start making calls to your deployed model via the REST API, Python, C#, or from the Studio. If your deployed model has a GPT-3.5 or GPT-4 model base, use the [Chat completions documentation](#), which has different request endpoints and variables required than for other base models.

All units complete:

[Complete module](#)