

## Converse

Service: Amazon Bedrock Runtime

Sends messages to the specified Amazon Bedrock model. Converse provides a consistent interface that works with all models that support messages. This allows you to write code once and use it with different models. If a model has unique inference parameters, you can also pass those unique parameters to the model.

Amazon Bedrock doesn't store any text, images, or documents that you provide as content. The data is only used to generate the response.

You can submit a prompt by including it in the `messages` field, specifying the `modelId` of a foundation model or inference profile to run inference on it, and including any other fields that are relevant to your use case.

You can also submit a prompt from Prompt management by specifying the ARN of the prompt version and including a map of variables to values in the `promptVariables` field. You can append more messages to the prompt by using the `messages` field. If you use a prompt from Prompt management, you can't include the following fields in the request: `additionalModelRequestFields`, `inferenceConfig`, `system`, or `toolConfig`. Instead, these fields must be defined through Prompt management. For more information, see [Test a prompt using Prompt management](#).

For information about the Converse API, see [Use the Converse API](#). To use a guardrail, see [Use a guardrail with the Converse API](#). To use a tool with a model, see [Tool use \(Function calling\)](#).

For example code, see [Converse API examples](#).

This operation requires permission for the `bedrock:InvokeModel` action.

### Important

To deny all inference access to resources that you specify in the `modelId` field, you need to deny access to the `bedrock:InvokeModel` and `bedrock:InvokeModelWithResponseStream` actions. Doing this also denies access to the resource through the base inference actions ([InvokeModel](#) and [InvokeModelWithResponseStream](#)). For more information see [Deny access for inference on specific models](#).

For troubleshooting some of the common errors you might encounter when using the Converse API, see [Troubleshooting Amazon Bedrock API Error Codes](#) in the Amazon Bedrock User Guide

## Request Syntax

POST /model/*modelId*/converse HTTP/1.1

Content-type: application/json

```
{
  "additionalModelRequestFields": JSON value,
  "additionalModelResponseFieldPaths": [ "string" ],
  "guardrailConfig": {
    "guardrailIdentifier": "string",
    "guardrailVersion": "string",
    "trace": "string"
  },
  "inferenceConfig": {
    "maxTokens": number,
    "stopSequences": [ "string" ],
    "temperature": number,
    "topP": number
  },
  "messages": [
    {
      "content": [
        { ... }
      ],
      "role": "string"
    }
  ],
  "performanceConfig": {
    "latency": "string"
  },
  "promptVariables": {
    "string" : { ... }
  },
  "requestMetadata": {
    "string" : "string"
  },
  "system": [
    { ... }
  ],
  "toolConfig": {
    "toolChoice": { ... },
```

```

    "tools": [
      { ... }
    ]
  }
}

```

## URI Request Parameters

The request uses the following URI parameters.

### modelId

Specifies the model or throughput with which to run inference, or the prompt resource to use in inference. The value depends on the resource that you use:

- If you use a base model, specify the model ID or its ARN. For a list of model IDs for base models, see [Amazon Bedrock base model IDs \(on-demand throughput\)](#) in the Amazon Bedrock User Guide.
- If you use an Amazon Bedrock Marketplace model, specify the ID or ARN of the marketplace endpoint that you created. For more information about Amazon Bedrock Marketplace and setting up an endpoint, see [Amazon Bedrock Marketplace](#) in the Amazon Bedrock User Guide.
- If you use an inference profile, specify the inference profile ID or its ARN. For a list of inference profile IDs, see [Supported Regions and models for cross-region inference](#) in the Amazon Bedrock User Guide.
- If you use a prompt created through [Prompt management](#), specify the ARN of the prompt version. For more information, see [Test a prompt using Prompt management](#).
- If you use a provisioned model, specify the ARN of the Provisioned Throughput. For more information, see [Run inference using a Provisioned Throughput](#) in the Amazon Bedrock User Guide.
- If you use a custom model, specify the ARN of the custom model deployment (for on-demand inference) or the ARN of your provisioned model (for Provisioned Throughput). For more information, see [Use a custom model in Amazon Bedrock](#) in the Amazon Bedrock User Guide.

Length Constraints: Minimum length of 1. Maximum length of 2048.

Pattern: `(arn:aws(-[^\:]+)?):bedrock:[a-z0-9-]{1,20}:((([0-9]{12}:custom-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}/[a-z0-9]{12}))|(:foundation-model/[a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.:]?[a-z0-9-]{1,63})))|([0-9]{12}:imported-model/[a-z0-9]{12})|([0-9]{12}:provisioned-model/[a-`

```
z0-9]{12})|([0-9]{12}:custom-model-deployment/[a-z0-9]{12})|([0-9]{12}:
(inference-profile|application-inference-profile)/[a-zA-Z0-9-:.]+))|
([a-z0-9-]{1,63}[.]{1}[a-z0-9-]{1,63}([.:]?[a-z0-9-]{1,63}))|(((0-9a-
zA-Z][_~]?)|([a-zA-Z0-9-:.]+)|(^arn:aws(-[:~]++)?:bedrock:[a-
z0-9-]{1,20}:([0-9]{12}:prompt/[0-9a-zA-Z]{10}(:?:[0-9]{1,5})?))$|
(^arn:aws:sagemaker:[a-z0-9-]+:([0-9]{12}:endpoint/[a-zA-Z0-9-]+)$|
(^arn:aws(-[:~]++)?:bedrock:([0-9a-z-]{1,20}):([0-9]{12}):default-)?
prompt-router/[a-zA-Z0-9-:.]+)$)
```

Required: Yes

## Request Body

The request accepts the following data in JSON format.

### additionalModelRequestFields

Additional inference parameters that the model supports, beyond the base set of inference parameters that Converse and ConverseStream support in the inferenceConfig field. For more information, see [Model parameters](#).

Type: JSON value

Required: No

### additionalModelResponseFieldPaths

Additional model parameters field paths to return in the response. Converse and ConverseStream return the requested fields as a JSON Pointer object in the additionalModelResponseFields field. The following is example JSON for additionalModelResponseFieldPaths.

```
[ "/stop_sequence" ]
```

For information about the JSON Pointer syntax, see the [Internet Engineering Task Force \(IETF\)](#) documentation.

Converse and ConverseStream reject an empty JSON Pointer or incorrectly structured JSON Pointer with a 400 error code. If the JSON Pointer is valid, but the requested field is not in the model response, it is ignored by Converse.

Type: Array of strings

Array Members: Minimum number of 0 items. Maximum number of 10 items.

Length Constraints: Minimum length of 1. Maximum length of 256.

Required: No

### **guardrailConfig**

Configuration information for a guardrail that you want to use in the request. If you include guardContent blocks in the content field in the messages field, the guardrail operates only on those messages. If you include no guardContent blocks, the guardrail operates on all messages in the request body and in any included prompt resource.

Type: [GuardrailConfiguration](#) object

Required: No

### **inferenceConfig**

Inference parameters to pass to the model. Converse and ConverseStream support a base set of inference parameters. If you need to pass additional parameters that the model supports, use the additionalModelRequestFields request field.

Type: [InferenceConfiguration](#) object

Required: No

### **messages**

The messages that you want to send to the model.

Type: Array of [Message](#) objects

Required: No

### **performanceConfig**

Model performance settings for the request.

Type: [PerformanceConfiguration](#) object

Required: No

## [promptVariables](#)

Contains a map of variables in a prompt from Prompt management to objects containing the values to fill in for them when running model invocation. This field is ignored if you don't specify a prompt resource in the `modelId` field.

Type: String to [PromptVariableValues](#) object map

Required: No

## [requestMetadata](#)

Key-value pairs that you can use to filter invocation logs.

Type: String to string map

Map Entries: Maximum number of 16 items.

Key Length Constraints: Minimum length of 1. Maximum length of 256.

Key Pattern: `[a-zA-Z0-9\s:_@$#=/+,-.]{1,256}`

Value Length Constraints: Minimum length of 0. Maximum length of 256.

Value Pattern: `[a-zA-Z0-9\s:_@$#=/+,-.]{0,256}`

Required: No

## [system](#)

A prompt that provides instructions or context to the model about the task it should perform, or the persona it should adopt during the conversation.

Type: Array of [SystemContentBlock](#) objects

Required: No

## [toolConfig](#)

Configuration information for the tools that the model can use when generating a response.

For information about models that support tool use, see [Supported models and model features](#).

Type: [ToolConfiguration](#) object

Required: No