

Filter

trainingPipelines

projects.locations.tuningJobs

Types

Annotation

AutomaticResources

BigQueryDestination

BigQuerySource

CsvSource

CustomJobSpec

DataItem

DedicatedResources

DestinationFeatureSetting

DiskSpec

EncryptionSpec

EnvVar

Event

Explanation

ExplanationSpec

FeatureSelector

FeatureValue

FeatureValueDestination

FeatureViewDataKey

FetchFeatureValuesResponse

GcsDestination

GcsSource

GenerateContentResponse

GenerationConfig

HarmCategory

JobState

LineageSubgraph

ListFeaturesResponse

MachineSpec

Measurement

ModelContainerSpec

ModelExplanation



NasTrial

NotebookIdleShutdownConfig

NotebookRuntimeType

Try [Gemini 1.5 models](#), the latest multimodal models in Vertex AI, and see what you can build with up to a 2M token context window.

Vertex AI > Documentation > Reference

Was this helpful?  

REST Resource: projects.locations.models

[Send feedback](#)

On this page

Resource: Model

ExportFormat

ExportableContent

DeploymentResourcesType

DeployedModelRef

ModelSourceInfo

ModelSourceType

OriginalModelInfo

...

Resource: Model

A trained machine learning Model.

JSON representation

```
{
  "name": string,
  "versionId": string,
  "versionAliases": [
    string
  ],
  "versionCreateTime": string,
  "versionUpdateTime": string,
  "displayName": string,
  "description": string,
  "versionDescription": string,
  "predictSchemata": {
    object (PredictSchemata)
  },
  "metadataSchemaUri": string,
  "metadata": value,
  "supportedExportFormats": [
    object (ModelContainerSpec)
  ],
  "artifactUri": string,
  "supportedDeploymentResourceTypes": [
    enum (DeploymentResourcesType)
  ],
  "supportedInputStorageFormats": [
    string
  ],
  "supportedOutputStorageFormats": [
    string
  ],
  "createTime": string,
  "updateTime": string,
  "deployedModels": [
    {
      object (DeployedModelRef)
    }
  ],
  "explanationSpec": {
    object (ExplanationSpec)
  },
  "etag": string,
  "labels": {
    string: string,
    ...
  },
  "encryptionSpec": {
```

cloud.google.com uses cookies to deliver and enhance the quality of its services and to analyze traffic. If you agree, cookies are also used to serve advertising and to personalize the content and advertisements that you see. [Learn more.](#)

Agree

No thanks

```
    object (ModelContainerSpec)
  ],
  "artifactUri": string,
  "supportedDeploymentResourceTypes": [
    enum (DeploymentResourcesType)
  ],
  "supportedInputStorageFormats": [
    string
  ],
  "supportedOutputStorageFormats": [
    string
  ],
  "createTime": string,
  "updateTime": string,
  "deployedModels": [
    {
      object (DeployedModelRef)
    }
  ],
  "explanationSpec": {
    object (ExplanationSpec)
  },
  "etag": string,
  "labels": {
    string: string,
    ...
  },
  "encryptionSpec": {
```

JSON representation

```
object (EncryptionSpec)
},
"modelSourceInfo": {
  object (ModelSourceInfo)
},
"originalModelInfo": {
  object (OriginalModelInfo)
},
"metadataArtifact": string,
"baseModelSource": {
  object (BaseModelSource)
},
"satisfiesPzs": boolean,
"satisfiesPzi": boolean
}
```

Fields

name	string The resource name of the Model.
versionId	string Output only. Immutable. The version ID of the model. A new version is committed when a new model version is uploaded or trained under an existing model id. It is an auto-incrementing decimal number in string representation.
versionAliases[]	string user provided version aliases so that a model version can be referenced via alias (i.e. projects/{project}/locations/{location}/models/{modelId}@{version_alias} instead of auto-generated version id (i.e. projects/{project}/locations/{location}/models/{modelId}@{versionId}). The format is [a-z][a-zA-Z0-9]{0,126}[a-z0-9] to distinguish from versionId. A default version alias will be created for the first version of the model, and there must be exactly one default version alias for a model.
versionCreateTime	string (Timestamp format) Output only. timestamp when this version was created. A timestamp in RFC3339 UTC "Zulu" format, with nanosecond resolution and up to nine fractional digits. Examples: "2014-10-02T15:01:23Z" and "2014-10-02T15:01:23.045123456Z".
versionUpdateTime	string (Timestamp format) Output only. timestamp when this version was most recently updated. A timestamp in RFC3339 UTC "Zulu" format, with nanosecond resolution and up to nine fractional digits. Examples: "2014-10-02T15:01:23Z" and "2014-10-02T15:01:23.045123456Z".
displayName	string Required. The display name of the Model. The name can be up to 128 characters long and can consist of any UTF-8 characters.
description	string The description of the Model.
versionDescription	string The description of this version.
predictSchema	object (PredictSchemata) The schemata that describe formats of the Model's predictions and explanations as given and returned via PredictionService.Predict and PredictionService.Explain .
metadataSchemaUri	string Immutable. Points to a YAML file stored on Google Cloud Storage describing additional information about the Model, that is specific to it. Unset if the Model does not have any additional information. The schema is defined as an OpenAPI 3.0.2 Schema Object . AutoML Models always have this field populated by Vertex AI, if no additional metadata is needed, this field is set to an empty string. Note: The URI given on output will be immutable and probably different, including the URI scheme, than the one given on input. The

Fields	
	output URI will point to a location where the user only has a read access.
metadata	<p>value (Value format)</p> <p>Immutable. An additional information about the Model; the schema of the metadata can be found in metadataSchema. Unset if the Model does not have any additional information.</p>
supportedExportFormats[]	<p>object (ExportFormat)</p> <p>Output only. The formats in which this Model may be exported. If empty, this Model is not available for export.</p>
trainingPipeline	<p>string</p> <p>Output only. The resource name of the TrainingPipeline that uploaded this Model, if any.</p>
pipelineJob	<p>string</p> <p>Optional. This field is populated if the model is produced by a pipeline job.</p>
containerSpec	<p>object (ModelContainerSpec)</p> <p>Input only. The specification of the container that is to be used when deploying this Model. The specification is ingested upon ModelService.UploadModel, and all binaries it contains are copied and stored internally by Vertex AI. Not required for AutoML Models.</p>
artifactUri	<p>string</p> <p>Immutable. The path to the directory containing the Model artifact and any of its supporting files. Not required for AutoML Models.</p>
supportedDeploymentResourceTypes[]	<p>enum (DeploymentResourcesType)</p> <p>Output only. When this Model is deployed, its prediction resources are described by the <code>prediction_resources</code> field of the Endpoint.deployed_models object. Because not all Models support all resource configuration types, the configuration types this Model supports are listed here. If no configuration types are listed, the Model cannot be deployed to an Endpoint and does not support online predictions (PredictionService.Predict or PredictionService.Explain). Such a Model can serve predictions by using a BatchPredictionJob, if it has at least one entry each in supportedInputStorageFormats and supportedOutputStorageFormats.</p>
supportedInputStorageFormats[]	<p>string</p> <p>Output only. The formats this Model supports in BatchPredictionJob.input_config. If PredictSchemata.instance_schema_uri exists, the instances should be given as per that schema.</p> <p>The possible formats are:</p> <ul style="list-style-type: none"> • <code>jsonl</code> The JSON Lines format, where each instance is a single line. uses GcsSource. • <code>csv</code> The CSV format, where each instance is a single comma-separated line. The first line in the file is the header, containing comma-separated field names. uses GcsSource. • <code>tf-record</code> The TFRecord format, where each instance is a single record in tfrecord syntax. uses GcsSource. • <code>tf-record-gzip</code> Similar to <code>tf-record</code>, but the file is gzipped. uses GcsSource. • <code>bigquery</code> Each instance is a single row in BigQuery. uses BigQuerySource. • <code>file-list</code> Each line of the file is the location of an instance to process, uses <code>gcsSource</code> field of the InputConfig object. <p>If this Model doesn't support any of these formats it means it cannot be used with a BatchPredictionJob. However, if it has supportedDeploymentResourceTypes, it could serve online predictions by using PredictionService.Predict or PredictionService.Explain.</p>
supportedOutputStorageFormats[]	<p>string</p> <p>Output only. The formats this Model supports in BatchPredictionJob.output_config. If both PredictSchemata.instance_schema_uri and PredictSchemata.prediction_schema_uri exist, the</p>

Fields	
	<p>predictions are returned together with their instances. In other words, the prediction has the original instance data first, followed by the actual prediction content (as per the schema).</p> <p>The possible formats are:</p> <ul style="list-style-type: none"> • <code>jsonl</code> The JSON Lines format, where each prediction is a single line. uses GcsDestination. • <code>csv</code> The CSV format, where each prediction is a single comma-separated line. The first line in the file is the header, containing comma-separated field names. uses GcsDestination. • <code>bigquery</code> Each prediction is a single row in a BigQuery table, uses BigQueryDestination. <p>If this Model doesn't support any of these formats it means it cannot be used with a BatchPredictionJob. However, if it has supportedDeploymentResourceTypes, it could serve online predictions by using PredictionService.Predict or PredictionService.Explain.</p>
<code>createTime</code>	<p>string (Timestamp format)</p> <p>Output only. timestamp when this Model was uploaded into Vertex AI.</p> <p>A timestamp in RFC3339 UTC "Zulu" format, with nanosecond resolution and up to nine fractional digits. Examples: "2014-10-02T15:01:23Z" and "2014-10-02T15:01:23.045123456Z".</p>
<code>updateTime</code>	<p>string (Timestamp format)</p> <p>Output only. timestamp when this Model was most recently updated.</p> <p>A timestamp in RFC3339 UTC "Zulu" format, with nanosecond resolution and up to nine fractional digits. Examples: "2014-10-02T15:01:23Z" and "2014-10-02T15:01:23.045123456Z".</p>
<code>deployedModels[]</code>	<p>object (DeployedModelRef)</p> <p>Output only. The pointers to DeployedModels created from this Model. Note that Model could have been deployed to Endpoints in different Locations.</p>
<code>explanationSpec</code>	<p>object (ExplanationSpec)</p> <p>The default explanation specification for this Model.</p> <p>The Model can be used for requesting explanation after being deployed if it is populated. The Model can be used for batch explanation if it is populated.</p> <p>All fields of the explanationSpec can be overridden by explanationSpec of DeployModelRequest.deployed_model, or explanationSpec of BatchPredictionJob.</p> <p>If the default explanation specification is not set for this Model, this Model can still be used for requesting explanation by setting explanationSpec of DeployModelRequest.deployed_model and for batch explanation by setting explanationSpec of BatchPredictionJob.</p>
<code>etag</code>	<p>string</p> <p>Used to perform consistent read-modify-write updates. If not set, a blind "overwrite" update happens.</p>
<code>labels</code>	<p>map (key: string, value: string)</p> <p>The labels with user-defined metadata to organize your Models.</p> <p>label keys and values can be no longer than 64 characters (Unicode codepoints), can only contain lowercase letters, numeric characters, underscores and dashes. International characters are allowed.</p> <p>See https://goo.gl/xmQnxf for more information and examples of labels.</p>
<code>encryptionSpec</code>	<p>object (EncryptionSpec)</p> <p>Customer-managed encryption key spec for a Model. If set, this Model and all sub-resources of this Model will be secured by this key.</p>
<code>modelSourceInfo</code>	<p>object (ModelSourceInfo)</p>

Fields	
fo	Output only. Source of a model. It can either be automl training pipeline, custom training pipeline, BigQuery ML, or saved and tuned from Genie or Model Garden.
originalModelInfo	object (OriginalModelInfo) Output only. If this Model is a copy of another Model, this contains info about the original.
metadataArtifact	string Output only. The resource name of the Artifact that was created in MetadataStore when creating the Model. The Artifact resource name pattern is projects/{project}/locations/{location}/metadataStores/{metadataStore}/artifacts/{artifact}.
baseModelSource	object (BaseModelSource) Optional. user input field to specify the base model source. Currently it only supports specifying the Model Garden models and Genie models.
satisfiesPzs	boolean Output only. reserved for future use.
satisfiesPzi	boolean Output only. reserved for future use.

ExportFormat

Represents export format supported by the Model. All formats export to Google Cloud Storage.

JSON representation
<pre>{ "id": string, "exportableContents": [enum (ExportableContent)] }</pre>

Fields	
id	string Output only. The ID of the export format. The possible format IDs are: <ul style="list-style-type: none"> tf-lite Used for Android mobile devices. edgetpu-tf-lite Used for Edge TPU devices. tf-saved-model A tensorflow model in SavedModel format. tf-js A TensorFlow.js model that can be used in the browser and in Node.js using JavaScript. core-ml Used for iOS mobile devices. custom-trained A Model that was uploaded or trained by custom code.
exportableContents[]	enum (ExportableContent) Output only. The content of this Model that may be exported.

ExportableContent

The Model content that can be exported.

Enums	
EXPORTABLE_CONTENT_UNSPECIFIED	Should not be used.
ARTIFACT	Model artifact and any of its supported files. Will be exported to the location specified by the artifactDestination field of the ExportModelRequest.output_config object.
IMAGE	The container image that is to be used when deploying this Model. Will be exported to the location specified by the imageDestination field of the

`ExportModelRequest.output_config` object.

DeploymentResourcesType

Identifies a type of Model's prediction resources.

Enums	
DEPLOYMENT_RESOURCES_TYPE_UNSPECIFIED	Should not be used.
DEDICATED_RESOURCES	Resources that are dedicated to the <code>DeployedModel</code> , and that need a higher degree of manual configuration.
AUTOMATIC_RESOURCES	Resources that to large degree are decided by Vertex AI, and require only a modest additional configuration.
SHARED_RESOURCES	Resources that can be shared by multiple <code>DeployedModels</code> . A pre-configured <code>DeploymentResourcePool</code> is required.

DeployedModelRef

Points to a `DeployedModel`.

JSON representation	
<pre>{ "endpoint": string, "deployedModelId": string }</pre>	
Fields	
endpoint	string Immutable. A resource name of an Endpoint.
deployedModelId	string Immutable. An ID of a <code>DeployedModel</code> in the above Endpoint.

ModelSourceInfo

Detail description of the source information of the model.

JSON representation	
<pre>{ "sourceType": enum (ModelSourceType) }</pre>	
Fields	
sourceType	enum (<code>ModelSourceType</code>) type of the model source.

ModelSourceType

Source of the model. Different from `objective` field, this `ModelSourceType` enum indicates the source from which the model was accessed or obtained, whereas the `objective` indicates the overall aim or function of this model.

Enums	
MODEL_SOURCE_TYPE_UNSPECIFIED	Should not be used.
AUTOML	The Model is uploaded by automl training pipeline.
CUSTOM	The Model is uploaded by user or custom training pipeline.
BQML	The Model is registered and sync'ed from BigQuery ML.
CUSTOM_TEXT_EMBEDDING	The Model is uploaded by text embedding finetuning pipeline.
MARKETPLACE	The Model is saved or tuned from Marketplace.

OriginalModelInfo

Contains information about the original Model if this Model is a copy.

JSON representation	
<pre>{ "model": string }</pre>	
Fields	
model	string Output only. The resource name of the Model this Model is a copy of, including the revision. Format: projects/{project}/locations/{location}/models/{modelId}@{versionId}

BaseModelSource

user input field to specify the base model source. Currently it only supports specifying the Model Garden models and Genie models.

JSON representation	
<pre>{ // Union field source can be only one of the following: "modelGardenSource": { object (ModelGardenSource) }, "genieSource": { object (GenieSource) } // End of list of possible types for union field source. }</pre>	
Fields	
Union field source.	
source can be only one of the following:	
modelGardenSource	object (ModelGardenSource) Source information of Model Garden models.
genieSource	object (GenieSource) Information about the base model of Genie models.

ModelGardenSource

Contains information about the source of the models generated from Model Garden.

JSON representation	
<pre>{ "publicModelName": string }</pre>	
Fields	
publicModelName	string Required. The model garden source model resource name.

GenieSource

Contains information about the source of the models generated from Generative AI Studio.

JSON representation	
<pre>{ "baseModelUri": string }</pre>	
Fields	
baseModelUri	string Required. The public base model URI.

Methods	
copy	Copies an already existing Vertex AI Model into the specified Location.
delete	Deletes a Model.
deleteVersion	Deletes a Model version.
export	Exports a trained, exportable Model to a location specified by the user.
get	Gets a Model.
getIamPolicy	Gets the access control policy for a resource.
list	Lists Models in a Location.
listVersions	Lists versions of the specified model.
mergeVersion...	Merges a set of aliases for a Model version.
patch	Updates a Model.
setIamPolicy	Sets the access control policy on the specified resource.
testIamPermi...	Returns permissions that a caller has on the specified resource.
updateExplan...	Incrementally update the dataset used for an examples model.
upload	Uploads a Model artifact into Vertex AI.

Was this helpful?




[Send feedback](#)

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 4.0 License](#), and code samples are licensed under the [Apache 2.0 License](#). For details, see the [Google Developers Site Policies](#). Java is a registered trademark of Oracle and/or its affiliates.

Last updated 2024-07-03 UTC.

Why Google	Products and	Solutions	Resources	Engage
Choosing Google Cloud	pricing	Infrastructure modernization	Google Cloud Affiliate Program	Contact sales
Trust and security	Google Cloud pricing	Databases	Google Cloud documentation	Find a Partner
Open cloud	Google Workspace pricing	Application modernization	Google Cloud quickstarts	Become a Partner
Multicloud	See all products	Smart analytics	Google Cloud Marketplace	Events
Global infrastructure		Artificial Intelligence	Learn about cloud computing	Podcasts
Customers and case studies		Security	Support	Developer Center
Analyst reports		Productivity & work transformation	Code samples	Press Corner
Whitepapers		Industry solutions	Cloud Architecture Center	Google Cloud on YouTube
Blog		DevOps solutions	Training	Google Cloud Tech on YouTube
		Small business solutions	Certifications	Follow on X
		See all solutions	Google for Developers	Join User Research
			Google Cloud for Startups	We're hiring. Join Google Cloud!
			System status	Google Cloud Community
			Release Notes	

[About Google](#) | [Privacy](#) | [Site terms](#) | [Google Cloud terms](#)

| [Manage cookies](#)  Our third decade of climate action: join us

Sign up for the Google Cloud newsletter

[Subscribe](#)

 [Language](#) ▾

