AI Platform

Fully managed, end-to-end platform for data science and machine learning.

Use AI Platform to train your machine learning models at scale, to host your trained model in the cloud, and to use your model to make predictions about new data.

* Managed services for hassle-free development
* Faster time to production with code-based and no-code tools
* Robust governance with interpretable models

### **Pricing**

AI Platform offers scalable, flexible pricing options to fit your project and budget.

AI Platform charges you for training your models and getting predictions. There is no charge for using AI Platform Vizier, AI Platform Notebooks, AI Platform Deep Learning Containers, AI Platform Deep Learning VM Image, or AI Platform Pipelines. However, you do pay for any Google Cloud resources you use with these products.

You can also use our [pricing calculator](https://cloud.google.com/products/calculator) to estimate the costs of running your workloads.

BENEFITS

For every skill level

Whether it's point-and-click data science using AutoML or advanced model optimization, AI Platform helps all users take their projects from ideation to deployment, quickly and seamlessly.

MLOps, simplified

Machine learning doesn’t stop at deployment. AI Platform makes it easy for developers, data scientists, and data engineers to streamline and scale their ML workflows.

Best of Google's AI

Take advantage of Google’s expertise in AI by infusing our cutting-edge technologies into your applications via tools on AI Platform like TPUs and TensorFlow.

KEY FEATURES

End-to-end machine learning life cycle

Prepare

Prepare and store your datasets with BigQuery and Cloud Storage, then use the built-in Data Labeling Service to label your training data for classification, object detection, entity extraction, and other objectives for image, video, tabular, and text data.

Build

Build best-in-class ML models without writing any code with AutoML's easy-to-use UI, or using your own code written in Notebooks, a managed Jupyter Notebook service. Use the latest open-source deep learning frameworks on Deep Learning VM Image or Deep Learning Containers. Then train your models with our fully managed Training service.

Validate

Validate your model with AI Explanations and What-If Tool, which help you understand your model's outputs, verify model behavior, identify bias, and find ways to improve your model and training data. Take model tuning a step further using Vizier, a black-box optimization service, to tune hyperparameters and optimize your model’s performance.

Deploy

Deploy your models at scale to get predictions in the cloud with Prediction, which hosts your model for online and batch prediction requests. You can also use AutoML Vision Edge to deploy your models at the edge and trigger real-time actions based on local data. TensorFlow Enterprise offers enterprise-grade support for TensorFlow instances.

MLOps

Manage your models, experiments, and end-to-end workflows with Pipelines by applying MLOps best practices with robust, repeatable pipelines. Continuous evaluation helps you monitor your models' performance, and provides continual feedback over time.

### **All your AI tools in one platform**

|  |  |
| --- | --- |
| AI Explanations | Understand how each feature in your input data contributed to model's outputs. |
| AutoML | Easily develop high-quality custom machine learning models without writing training routines. Powered by Google’s state-of-the-art transfer learning and hyperparameter search technology. |
| Continuous evaluation | Obtain metrics about the performance of your models in production. Compare predictions with ground truth labels to gain continual feedback and optimize model performance over time. |
| Data Labeling Service | Get highly accurate labels from human labelers for better machine learning models. |
| Deep Learning Containers | Quickly build and deploy models in a portable and consistent environment for all your AI applications. |
| Deep Learning VM Image | Instantiate a VM image containing the most popular AI frameworks on a Compute Engine instance without worrying about software compatibility. |
| Neural Architecture Search | Build application-specific models and improve existing model architectures with an automated service. Powered by Google’s leading AI research, users can design models that are optimized for latency, accuracy, power consumption, and more. |
| Notebooks | Create, manage, and connect to VMs with JupyterLab, the standard data scientist workbench. VMs come pre-installed deep learning frameworks and libraries. |
| Pipelines | Implement MLOps by orchestrating the steps in your ML workflow as a pipeline without the difficulty of setting up Kubeflow Pipelines with TensorFlow Extended (TFX). |
| Prediction | Easily deploy your models to managed, scalable endpoints for online or batch predictions. |
| TensorFlow Enterprise | Easily develop and deploy TensorFlow models on Google Cloud with enterprise-grade support and cloud scale performance. |
| Training | Train any models in any framework on any hardware, from single machines to large clusters with multiple accelerators. |
| Vizier | Optimize your model's output by intelligently tuning hyperparameters. |
| What-If Tool | Visualize your datasets and probe your model to better understand its behavior with an interactive visual interface. |

ML workflow

As the diagram indicates, you can use AI Platform to manage the following stages in the ML workflow:

* Train an ML model on your data:
  + Train model
  + Evaluate model accuracy
  + Tune hyperparameters
* Deploy your trained model.
* Send prediction requests to your model:
  + Online prediction
  + Batch prediction *(for TensorFlow only)*
* Monitor the predictions on an ongoing basis.
* Manage your models and model versions.

Components of AI Platform

Training service

The AI Platform training service allows you to train models using a wide range of different customization options.

You can select many different machine types to power your training jobs, enable distributed training, use hyperparameter tuning, and accelerate with GPUs and TPUs.

You can also select different ways to customize your training application. You can submit your input data for AI Platform to train using a built-in algorithm (beta). If the built-in algorithms do not fit your use case, you can submit your own training application to run on AI Platform, or build a custom container with your training application and its dependencies to run on AI Platform.

Prediction service

The AI Platform prediction service allows you to serve predictions based on a trained model, whether or not the model was trained on AI Platform.

Notebooks

AI Platform Notebooks enables you to create and manage virtual machine (VM) instances that are pre-packaged with JupyterLab.

AI Platform Notebooks instances have a pre-installed suite of deep learning packages, including support for the TensorFlow and PyTorch frameworks. You can configure either CPU-only or GPU-enabled instances, to best suit your needs.

Your notebook instances are protected by Google Cloud authentication and authorization, and are available using a notebook instance URL. Notebook instances also integrate with GitHub so that you can easily sync your notebook with a GitHub repository.

Data labeling service

AI Platform Data Labeling Service (beta) lets you request human labeling for a dataset that you plan to use to train a custom machine learning model. You can submit a request to label your video, image, or text data.

To submit a labeling request, you provide a representative sample of labeled data, specify all the possible labels for your dataset, and provide some instructions for how to apply those labels. The human labelers follow your instructions, and when the labeling request is complete, you get your annotated dataset that you can use to train a machine learning model.

Deep learning VM image

AI Platform Deep Learning VM Image lets you choose from a set of Debian 9-based Compute Engine virtual machine images optimized for data science and machine learning tasks. All images come with key ML frameworks and tools pre-installed, and can be used out of the box on instances with GPUs to accelerate your data processing tasks.

## Tools to interact with AI Platform

Google Cloud Console

You can deploy models to the cloud and manage your models, versions, and jobs on the Cloud Console. This option gives you a user interface for working with your machine learning resources. As part of Google Cloud, your AI Platform resources are connected to useful tools like Cloud Logging and Cloud Monitoring.

The gcloud command-line tool

You can manage your models and versions, submit jobs, and accomplish other AI Platform tasks at the command line with the gcloud ai-platform command-line tool.

We recommend gcloud commands for most AI Platform tasks, and the REST API (see below) for online predictions.

REST API

The AI Platform REST API provides RESTful services for managing jobs, models, and versions, and for making predictions with hosted models on Google Cloud.

You can use the Google APIs Client Library for Python to access the APIs. When using the client library, you use Python representations of the resources and objects used by the API. This is easier and requires less code than working directly with HTTP requests.

We recommend the REST API for serving online predictions in particular.