

[AWS](#) > [Documentation](#) > [Amazon SageMaker](#) > **Developer Guide**

# Use Amazon SageMaker Ground Truth to Label Data

[PDF \(/pdfs/sagemaker/latest/dg/sagemaker-dg.pdf#sms\)](#)

[RSS \(amazon-sagemaker-release-notes.rss\)](#)

To train a machine learning model, you need a large, high-quality, labeled dataset. Ground Truth helps you build high-quality training datasets for your machine learning models. With Ground Truth, you can use workers from either Amazon Mechanical Turk, a vendor company that you choose, or an internal, private workforce along with machine learning to enable you to create a labeled dataset. You can use the labeled dataset output from Ground Truth to train your own models. You can also use the output as a training dataset for an Amazon SageMaker model.

Depending on your ML application, you can choose from one of the Ground Truth built-in task types to have workers generate specific types of labels for your data. You can also build a custom labeling workflow to provide your own UI and tools to workers labeling your data. To learn more about the Ground Truth built in task types, see [Built-in Task Types \(/sms-task-types.html\)](#) . To learn how to create a custom labeling workflow, see [Creating Custom Labeling Workflows \(/sms-custom-templates.html\)](#) .

In order to automate labeling your training dataset, you can optionally use *automated data labeling*, a Ground Truth process that uses machine learning to decide which data needs to be labeled by humans. Automated data labeling may reduce the labeling time and manual effort required. For more information,

see [Automate Data Labeling \(./sms-automated-labeling.html\)](#) . To create a custom labeling workflow, see [Creating Custom Labeling Workflows \(./sms-custom-templates.html\)](#) .

Use either pre-built or custom tools to assign the labeling tasks for your training dataset. A *labeling UI template* is a webpage that Ground Truth uses to present tasks and instructions to your workers. The SageMaker console provides built-in templates for labeling data. You can use these templates to get started , or you can build your own tasks and instructions by using our HTML 2.0 components. For more information, see [Creating Custom Labeling Workflows \(./sms-custom-templates.html\)](#) .

Use the workforce of your choice to label your dataset. You can choose your workforce from:

- The Amazon Mechanical Turk workforce of over 500,000 independent contractors worldwide.
- A private workforce that you create from your employees or contractors for handling data within your organization.
- A vendor company that you can find in the AWS Marketplace that specializes in data labeling services.

For more information, see [Create and Manage Workforces \(./sms-workforce-management.html\)](#) .

You store your datasets in Amazon S3 buckets. The buckets contain three things: The data to be labeled, an input manifest file that Ground Truth uses to read the data files, and an output manifest file. The output file contains the results of the labeling job. For more information, see [Use Input and Output Data \(./sms-data.html\)](#) .

Events from your labeling jobs appear in Amazon CloudWatch under the `/aws/sagemaker/LabelingJobs` group. CloudWatch uses the labeling job name as the name for the log stream.

---

## Are You a First-time User of Ground Truth?

If you are a first-time user of Ground Truth, we recommend that you do the following:

1. **Read [Getting started \(./sms-getting-started.html\)](#)**—This section walks you through setting up your first Ground Truth labeling job.
  2. **Explore other topics**—Depending on your needs, do the following:
    - **Explore built-in task types**— Use built-in task types to streamline the process of creating a labeling job. See [Built-in Task Types \(./sms-task-types.html\)](#) to learn more about Ground Truth built-in task types.
    - **Manage your labeling workforce**—Create new work teams and manage your existing workforce. For more information, see [Create and Manage Workforces \(./sms-workforce-management.html\)](#) .
    - **Learn about streaming labeling jobs**— Create a streaming labeling job and send new dataset objects to workers in real time using a perpetually running labeling job. Workers continuously receive new data objects to label as long as the labeling job is active and new objects are being sent to it. To learn more, see [Ground Truth Streaming Labeling Jobs \(./sms-streaming-labeling-job.html\)](#) .
  3. **See the [Reference \(https://docs.aws.amazon.com/sagemaker/latest/APIReference/API\\_Reference.html\)](https://docs.aws.amazon.com/sagemaker/latest/APIReference/API_Reference.html)**  
—This section describes operations to automate Ground Truth operations.
- 

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.