# Review: Google Cloud AI lights up machine learning

### **Google Cloud AI and Machine Learning Platform is missing some pieces, and much is still in beta, but its scope and quality are second to none.**

Google has one of the largest machine learning stacks in the industry, currently centering on its Google Cloud AI and Machine Learning Platform. Google spun out TensorFlow as open source years ago, but [TensorFlow](https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html) is still the most mature and widely cited [deep learning framework](https://www.infoworld.com/article/3163525/the-best-machine-learning-and-deep-learning-libraries.html). Similarly, Google spun out Kubernetes as open source years ago, but it is still the dominant container management system.

Google is one of the top sources of tools and infrastructure for developers, data scientists, and machine learning experts, but historically Google AI hasn’t been all that attractive to business analysts who lack serious data science or programming backgrounds. That’s starting to change.

The Google Cloud AI and Machine Learning Platform includes AI building blocks, the AI platform and accelerators, and AI solutions. The AI solutions are fairly new and aimed at business managers rather than data scientists. They may include consulting from Google or its partners.

The AI building blocks, which are pre-trained but customizable, can be used without intimate knowledge of programming or data science. Nevertheless, they are often used by skilled data scientists for pragmatic reasons, essentially to get stuff done without extensive model training.

The AI platform and accelerators are generally for serious data scientists, and require coding skill, knowledge of data preparation techniques, and lots of training time. I recommend going there only after trying the relevant building blocks.

There are still some missing links in Google Cloud’s AI offerings, especially in data preparation. The closest thing Google Cloud has to a data import and conditioning service is the third-party Cloud Dataprep by Trifacta; I tried it a year ago and was underwhelmed. The feature engineering built into Cloud AutoML Tables is promising, however, and it would be useful to have that sort of service available for other scenarios.

The seamy underside of AI has to do with ethics and responsibility (or the lack thereof), along with persistent model biases (often because of biased data used for training). Google published its AI Principles in 2018. It’s a work in progress, but it’s a basis for guidance as discussed in a [recent blog post](https://cloud.google.com/blog/products/ai-machine-learning/taking-care-of-business-with-responsible-ai) on Responsible AI.

There is lots of competition in the AI market (over a dozen vendors), and lots of competition in the public cloud market (over half-a-dozen credible vendors). To do the comparisons justice, I’d have to write an article at least five times as long as this one, so as much as I hate leaving them out, I’ll have to omit most product comparisons. For the top obvious comparison, I can summarize: AWS does most of what Google does, and is also very good, but generally charges higher prices.

## Google Cloud AI Building Blocks

Google Cloud AI Building Blocks are easy-to-use components that you can incorporate into your own applications to add sight, language, conversation, and structured data. Many of the AI building blocks are pre-trained neural networks, but can be customized with transfer learning and neural network search if they don’t serve your needs out of the box. AutoML Tables is a little different, in that it automates the process a data scientist would use to find the best machine learning model for a tabular data set.

### AutoML

The Google Cloud AutoML services provide customized deep neural networks for language pair translation, text classification, object detection, image classification, and video object classification and tracking. They require tagged data for training, but don’t require significant knowledge of deep learning, transfer learning, or programming.

Google Cloud AutoML customizes Google’s battle-tested, high-accuracy deep neural networks for your tagged data. Rather than starting from scratch when training models from your data, AutoML implements automatic deep transfer learning (meaning that it starts from an existing deep neural network trained on other data) and neural architecture search (meaning that it finds the right combination of extra network layers) for language pair translation and the other services listed above.

In each area, Google already has one or more pre-trained services based on deep neural networks and huge sets of labeled data. These may well work for your data unmodified, and you should test that to save yourself time and money. If they don’t do what you need, Google Cloud AutoML helps you to create a model that does, without requiring that you know how to perform transfer learning or how to design neural networks.

Transfer learning offers two big advantages over training a neural network from scratch. First, it requires a lot less data for training, since most of the layers of the network are already well trained. Second, it trains a lot faster, since it’s only optimizing the final layers.

While the Google Cloud AutoML services used to be presented together as a package, they are now listed with their base pre-trained services. What most other companies call AutoML is performed by Google Cloud AutoML Tables.

### AutoML Tables

The usual data science process for many regression and classification problems is to create a table of data for training, clean and condition the data, perform feature engineering, and try to train all of the appropriate models on the transformed table, including a step to optimize the best models’ hyperparameters. Google Cloud AutoML Tables can perform this entire process automatically once you manually identify the target field.

AutoML Tables automatically searches through Google’s model zoo for structured data to find the best model for your needs, ranging from linear/logistic regression models for simpler data sets to advanced deep, ensemble, and architecture-search methods for larger, more complex ones. It automates feature engineering on a wide range of tabular data primitives — such as numbers, classes, strings, timestamps, and lists — and helps you detect and take care of missing values, outliers, and other common data issues.

Its codeless interface guides you through the full end-to-end machine learning lifecycle, making it easy for anyone on your team to build models and reliably incorporate them into broader applications. AutoML Tables provides extensive input data and model behavior explainability features, along with guardrails to prevent common mistakes. AutoML Tables is also available in API and notebook environments.

AutoML Tables competes with [Driverless AI](https://www.infoworld.com/article/3236048/machine-learning/review-h2oai-automates-machine-learning.html) and several other AutoML implementations and frameworks.

### Vision API

The Google Cloud Vision API is a pre-trained machine learning service for categorizing images and extracting various features. It can classify images into thousands of pre-trained categories, ranging from generic objects and animals found in the image (such as a cat), to general conditions (for example, dusk), to specific landmarks (Eiffel Tower, Grand Canyon), and identify general properties of the image, such as its dominant colors. It can isolate areas that are faces, then apply geometric (facial orientation and landmarks) and emotional analyses to the faces, although it does not recognize faces as belonging to specific people, except for celebrities (which requires a special usage license). Vision API uses OCR to detect text within images in more than 50 languages and various file types. It can also identify product logos, and detect adult, violent, and medical content.

[Read the full review of Google Cloud Machine Learning APIs](https://www.infoworld.com/article/3100454/first-look-google-cloud-machine-learning-soars.html)

### Video Intelligence API

The Google Cloud Video Intelligence API automatically recognizes more than 20,000 objects, places, and actions in stored and streaming video. It also distinguishes scene changes and extracts rich metadata at the video, shot, or frame level. It additionally performs text detection and extraction using OCR, detects explicit content, automates closed captioning and subtitles, recognizes logos, and detects faces, persons, and poses.

Google recommends the Video Intelligence API for extracting metadata to index, organize, and search your video content. It can transcribe videos and generate closed captions, as well as flag and filter inappropriate content, all more cost-effectively than human transcribers. Use cases include content moderation, content recommendations, media archives, and contextual advertisements.

### Natural Language API

[Natural language processing](https://www.infoworld.com/article/3398696/what-is-natural-language-processing-ai-for-speech-and-text.html) (NLP) is a big part of the “secret sauce” that makes input to Google Search and the Google Assistant work well. The Google Cloud Natural Language API exposes that same technology to your programs. It can perform syntax analysis (see the image below), entity extraction, sentiment analysis, and content classification, in 10 languages. You may specify the language if you know it; otherwise, the API will attempt to auto-detect the language. A separate API, currently available for early access on request, specializes in healthcare-related content.

[Read the full review of Google Cloud Machine Learning APIs](https://www.infoworld.com/article/3100454/first-look-google-cloud-machine-learning-soars.html)

### Translation

The Google Cloud Translation API can translate over a hundred language pairs, can auto-detect the source language if you don’t specify it, and comes in three flavors: Basic, Advanced, and Media Translation. The Advanced Translation API supports a glossary, batch translation, and the use of custom models. The Basic Translation API is essentially what is used by the consumer Google Translate interface. AutoML Translation allows you to train custom models using transfer learning.

The Media Translation API translates content directly from audio (speech), either audio files or streams, in 12 languages, and automatically generates punctuation. There are separate models for video and phone call audio.