

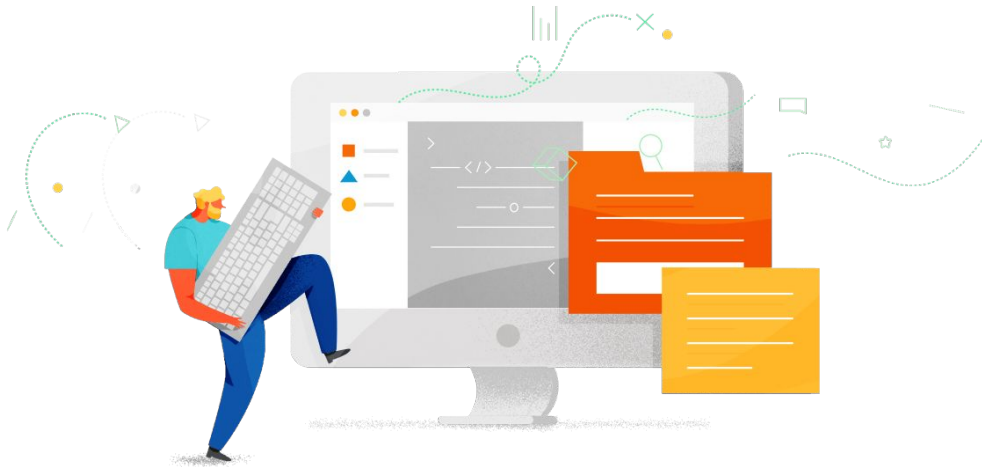
AI Platform Unified Google Cloud

Ana López-Mancisidor (analopezm@google.com)

Teresa Muñoz-Cruzado (teresamunoz@google.com)

Vanessa Peinó (vpeino@google.com)

Agenda



- 1 ¿Por qué Google en IA?
- 2 Enfoque IA en GCP
- 3 CCAI & Document AI
- 4 ML & APIs
- 5 AI Platform Unified

¿Por qué Google en IA?

Nuestras fortalezas en IA / ML



La organización de investigación de IA más grande del mundo



Contribuciones a la comunidad de desarrolladores y de código abierto



Calidad e innovación en la producción (consumidor)



Mejores prácticas de ingeniería para empresas y MLOps

Ejemplos de ML de Google en producción



Search

Search Ranking
Speech Recognition



Gmail

Smart Reply
Spam Classification



Photos

Photos Search



Translate

Text, Graphic and
Speech Translations



Android

Keyboard and
Speech Input



Drive

Intelligence in Apps



YouTube

Video Recommendations
Better Thumbnails



Data Center Power Usage

Reduced cooling
energy 40%



Self Driving Car

More than 1.5MM miles
driven



Play

App Recommendations
Game Developer Experience



Chrome

Search by Image



Maps

Street View Image
Parsing Local Search



Ads

Richer Text Ads
Automated Bidding

Nuestras fortalezas en IA / ML



**Accesible
via API**

Portabilidad

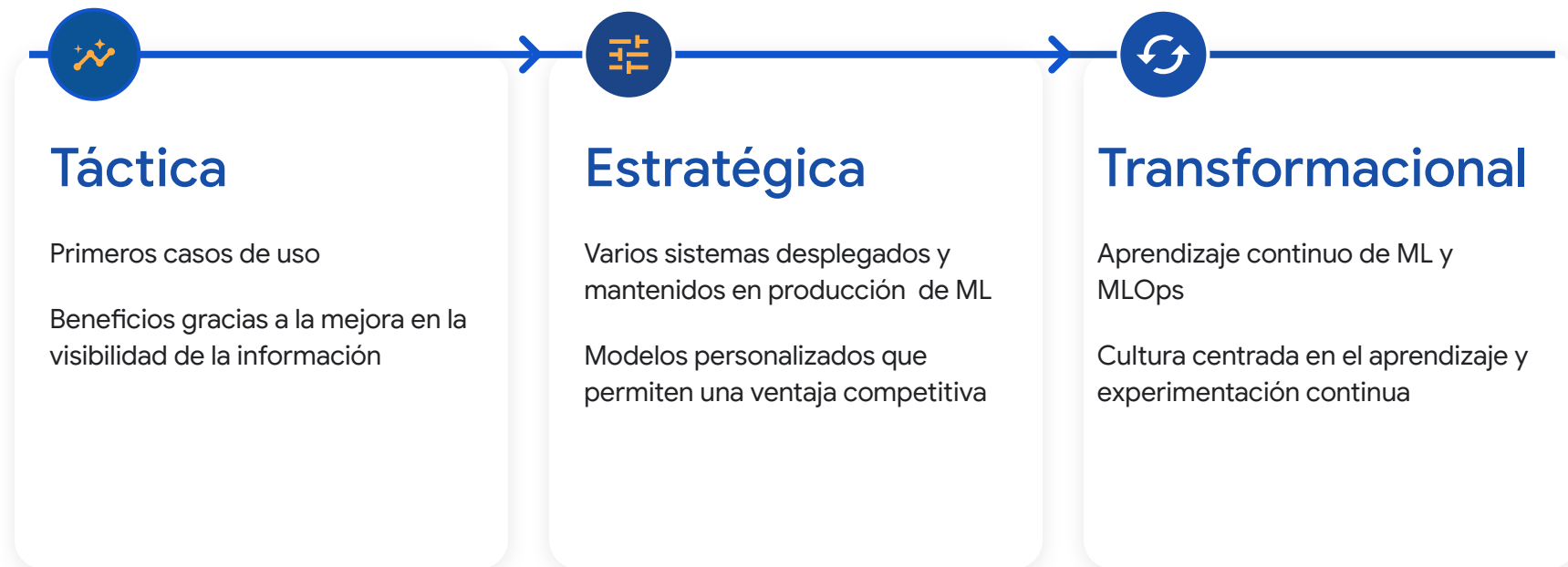
(TensorFlow,
Kubeflow)

**Enterprise
Ready**

Seguridad

(los datos de los clientes son de los clientes, y Google no usará dicha información para mejorar nuestros modelos ya entrenados o los modelos de otros clientes)

Camino hacia la madurez en la IA



Enfoque IA en GCP

Inteligencia Artificial en Google Cloud

Construye tus propios modelos:
tus datos + tu modelo



Cloud ML Engine



Kubeflow

Modelos

Personalizados:
tus datos +
nuestro modelo



Auto ML



Dialogflow

Invoca nuestras APIs:

nuestros datos + nuestros modelos



[Cloud Vision API](#)



[Cloud Speech API](#)



Cloud Jobs API



[Cloud Translation API](#)



[Cloud Natural Language API](#)



[Cloud Video Intelligence API](#)



Cloud Speech Synthesis API



[Data Loss Prevention API](#)

Cloud AI portfolio



Contact Center AI

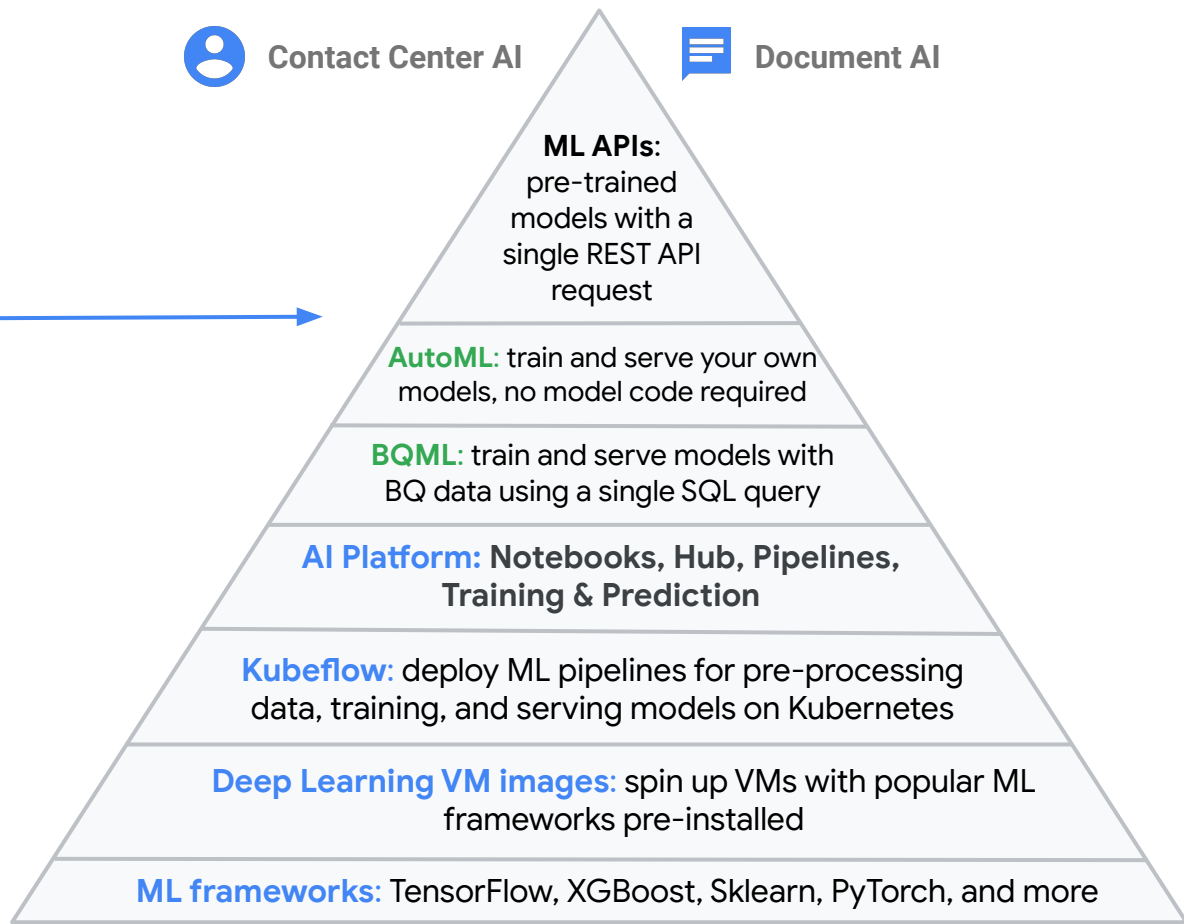


Document AI

Desarrolladores
de aplicaciones



Modelos
personalizados





CCAI

What is **Contact Center AI**?



Automates basic chat and
voice interactions

Dialogflow



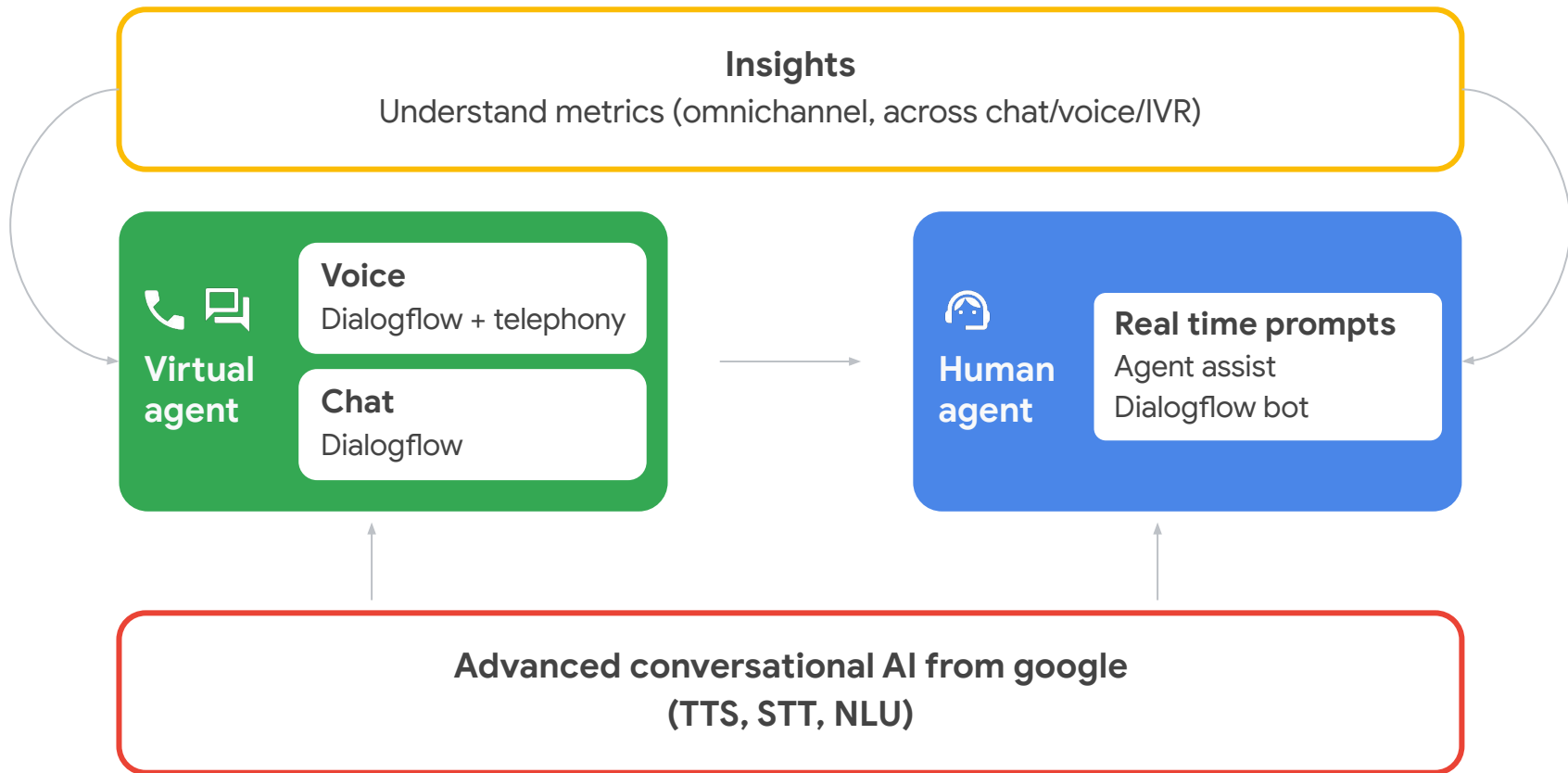
Makes human agents
more effective

Agent Assist



Unlocks insights about
call drivers

CCAI Insights



**Enabling a
conversation that
is close to human**

Understand

- Speech-to-text

Talk

- Text-to-speech

Interact

- Dialogflow
- Document assist
- Sentiment analysis

Understand

Using the most advanced deep-learning neural network algorithms, Google Cloud's **Speech-to-text** performs speech recognition with **unparalleled accuracy**.

10x

Larger vocabulary than the entire Oxford English Dictionary

120+

languages covered with more coming

~5000

distinct phrase hints can be provided at real-time to customize the recognizer

~4

languages can be auto-detected

Formatting

Real-life, context-specific formatting (e.g., \$ or phone # format) automatically added

Real-time

Returns text as it's recognized and can analyze short & long-form audio

Talk

With industry-leading voice quality, Google Cloud's **Text-to-speech** enables a natural, conversational manner across channels.

180+

Voices you can choose from

92

Wavenet voices

21

Languages covered with more coming

Wavenet

Exclusive multilingual access to DeepMind's WaveNet technology, which offer the most human-like voices

REST/GRPC

Seamless integration with an REST or GRPC capable application

Pronunciation

Allows pauses, numbers, date & time formatting, and other nuances

Standard



WaveNet



Try Text-to-Speech
with any text!

Try seamless, immediate,
speech-text-speech translation!

Interact

Reaching +1Mn developers, **Dialogflow** is the emerging standard for building natural and rich conversational experiences across multiple channels.

1M+

developers on the platform

32

Languages and variances with more coming

17

Single Click integrations with popular chat and messaging platforms

2000

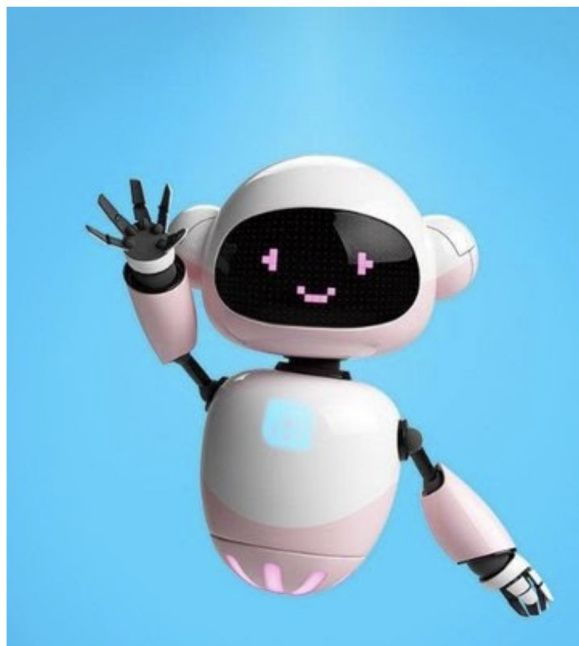
Intents. Speak to us if you need more.

REST/GRPC

HIPAA and PCI Compliant and similar SLO than other Cloud products

Flexible

Platform agnostic built on GCP, with Cloud Support and SLA available. 7 SDKS to integrate in your applications (python, node, java, Go, etc)



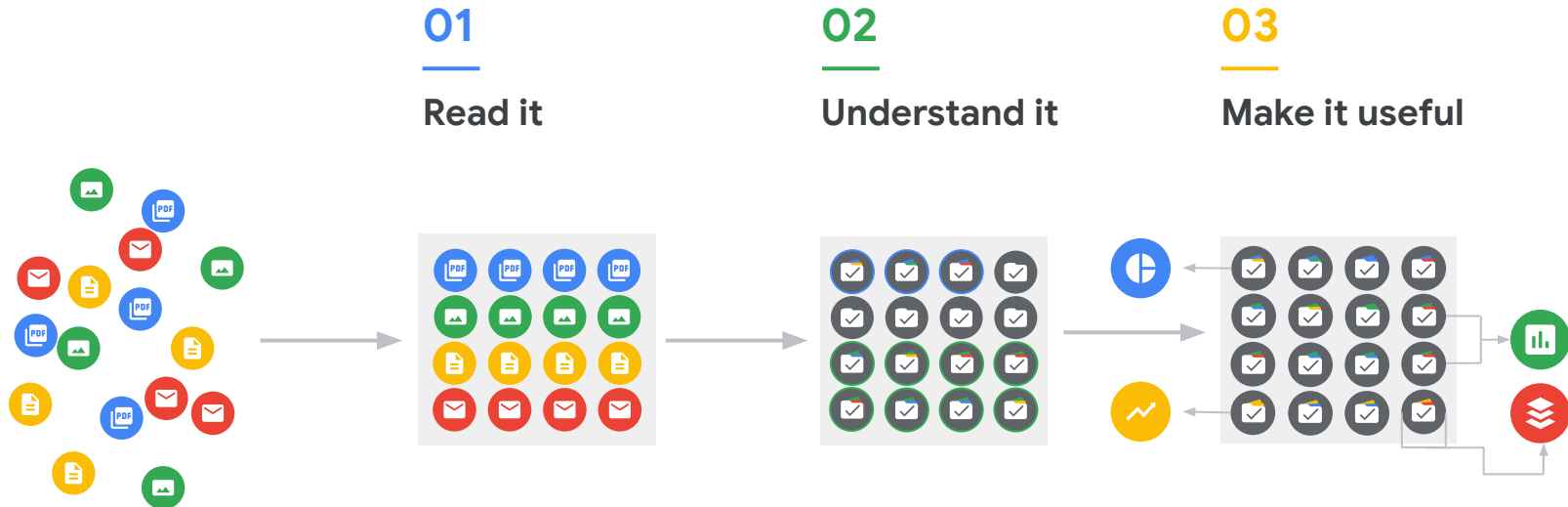
Soy COVA, tu agente experto en vacunación





Document AI

Document AI **extracts** & **classifies** information from unstructured documents



Life of a document

INGEST & FILTER

Document is read in from
Cloud Storage.

Unstructured
document, multiple
formats and languages



CATEGORIZE CONTENT

Categorize patent's content
using NLP model.

DETECT DIAGRAMS

Identify diagram and
corresponding x and y
coordinates.

OCR

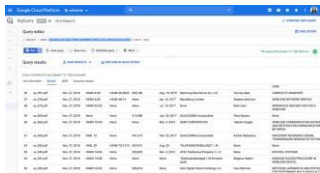
Extract out raw text into
json format for
downstream NLP process.

EXTRACT ENTITIES

Identify named entities in
the raw text.

STORAGE

Write out and store
results from the pipeline
into BigQuery



Document ID	Text	Entities	Diagram Coordinates
1	United States Patent	United States Patent	
2	Patent No. 1,234,567	Patent No. 1,234,567	
3	Abstract	Abstract	
4	Background	Background	
5	Summary	Summary	
6	Claims	Claims	
7	Description	Description	
8	References	References	
9	Figures	Figures	
10	Conclusion	Conclusion	

Demo:

<https://cloud.google.com/solutions/document-ai>

What is Document AI?

DocAI turns
unstructured content
into structured data.



Document image



```
{Class: /us/gov/ID/Driverslicense  
State: Ohio  
Name: Jane Doe  
Issued: 7/2/2018  
Expires: 1/21/2021 }
```

Structured data

118n: English, French, Dutch, German, Spanish; 150+ more languages by March 2021

Schema	Value
carrier	UPS
invoice_id	36139
payment_terms	2% 10 Net 30
ship_to_name	Graphic Packaging Int'l
receiver_name	Graphic Packaging Int'l Inc
invoice_date	9/7/2018
ship_to_address	1421 N. Pitcher Kalamazoo, MI 49007
amount_due	\$2,525.62
receiver_address	PO Box 35800 West Monroe, LA 71294-5800
total_amount	\$2,525.62
supplier_address	115 Progress Drive Manchester, CT 06042
line_item/product_code	BG 101
line_item/product_code	BG
line_item/quantity	2
line_item/quantity	20
line_item/quantity	3
line_item/product_code	BG 6507.14021
line_item/unit_price	222.59



ML APIs

Machine Learning APIs: Ready to Go



Cloud
Vision API



Cloud
Speech API



Cloud
Translation API

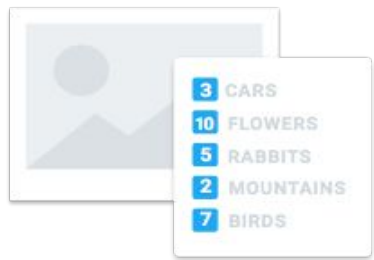


Cloud Natural
Language API



Cloud Video
Intelligence API

Vision API



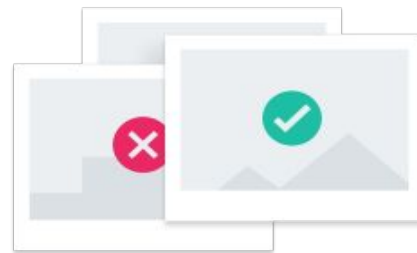
Object recognition



Logo and facial analysis

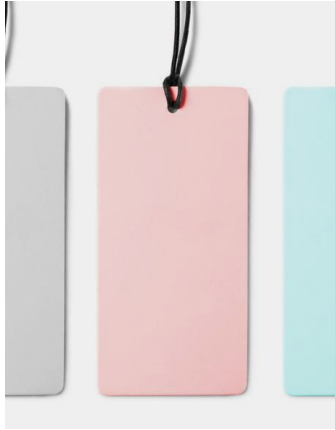


Text extractions



Detect inappropriate
content

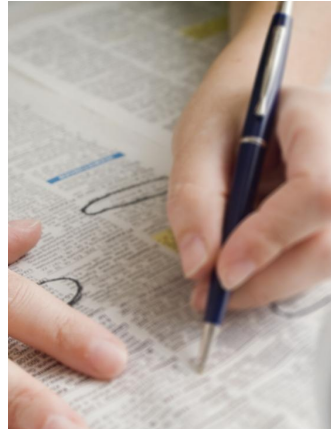
Natural Language API



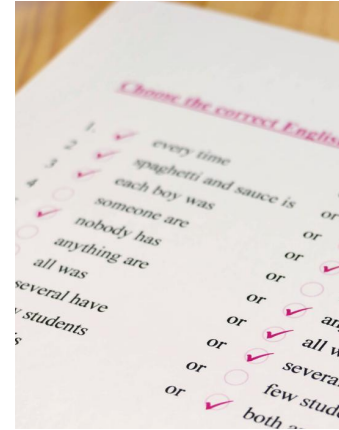
Classify content



Detect sentiment



Extract entities



Analyze syntax

Demo time!

<https://cloud.google.com/vision/docs/drag-and-drop>

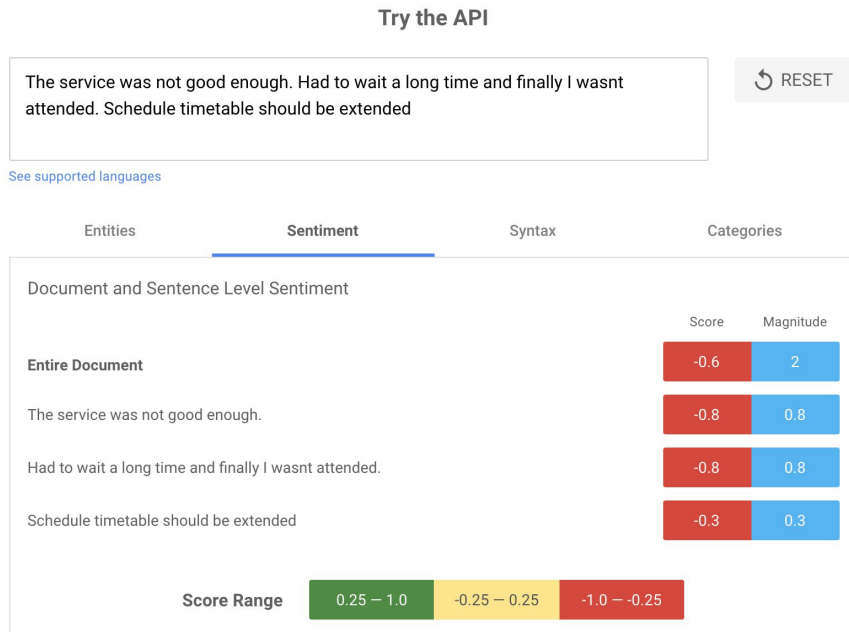


Natural Language API: Sentiment Analysis

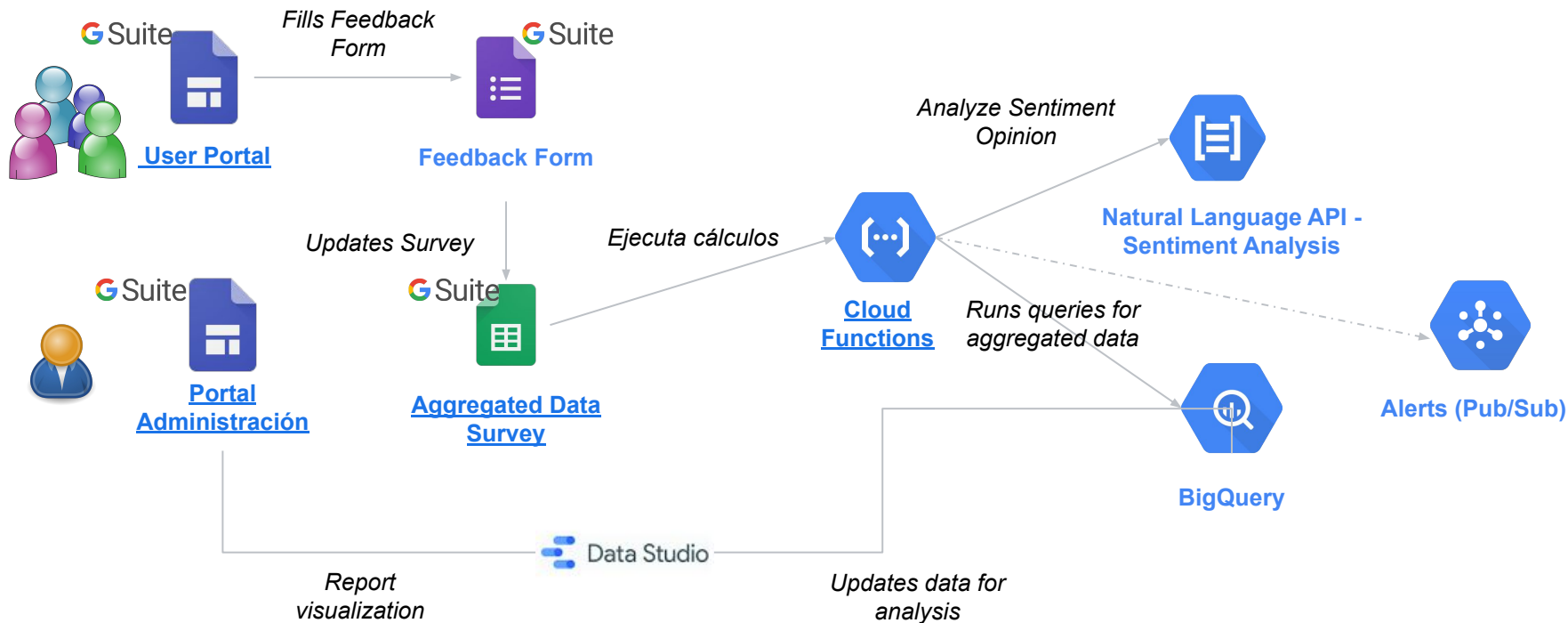
Inspects the given text and identifies the prevailing emotional opinion within the text, especially to determine a writer's attitude as positive, negative, or neutral.

- **Score:** positive, negative, neutral
- **Magnitude:** indicates the overall strength of emotion (both positive and negative) within the given text

Sentiment	Sample Values
Clearly Positive*	"score": 0.8, "magnitude": 3.0
Clearly Negative*	"score": -0.6, "magnitude": 4.0
Neutral	"score": 0.1, "magnitude": 0.0
Mixed	"score": 0.0, "magnitude": 4.0



Demo: Analyze the sentiment of user opinions with Google Cloud AI



ML Frameworks:

Total Control



TensorFlow



AI Platform



Spark ML



Cloud Dataproc

AutoML:

Bring Your Own Data
(We Do the Rest)



Machine Learning APIs:

Ready to Go



Cloud
Vision API



Cloud
Speech API



Cloud
Translation API



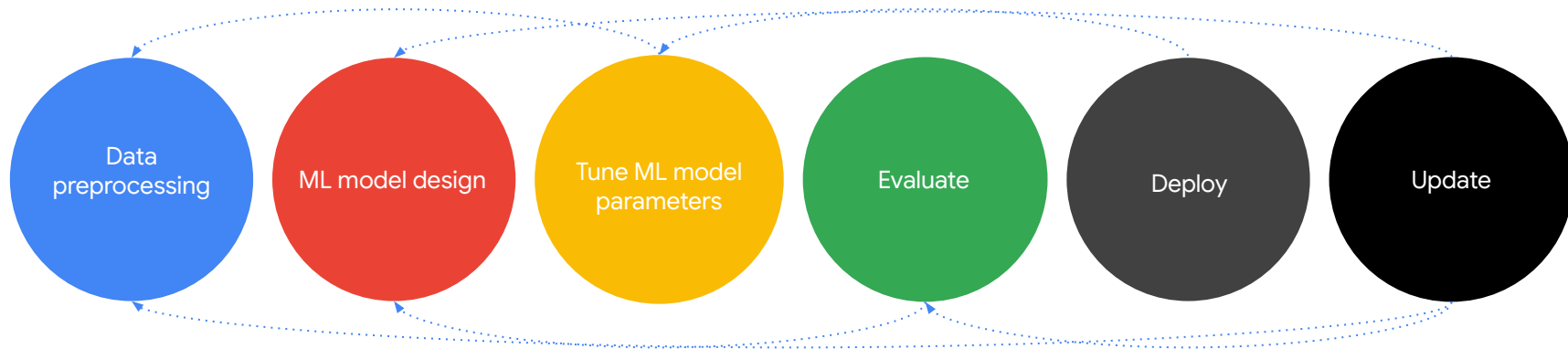
Cloud Natural
Language API

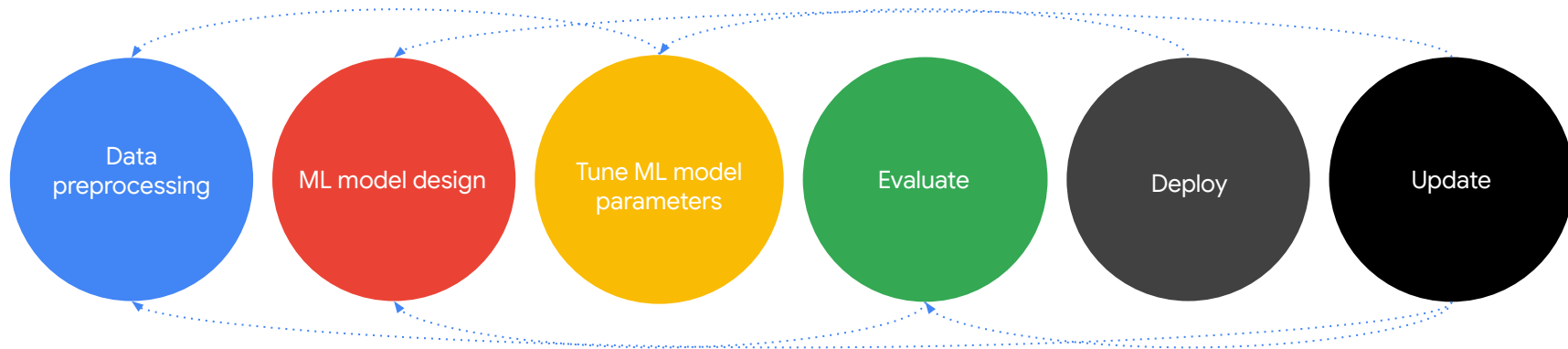


Cloud Video
Intelligence API



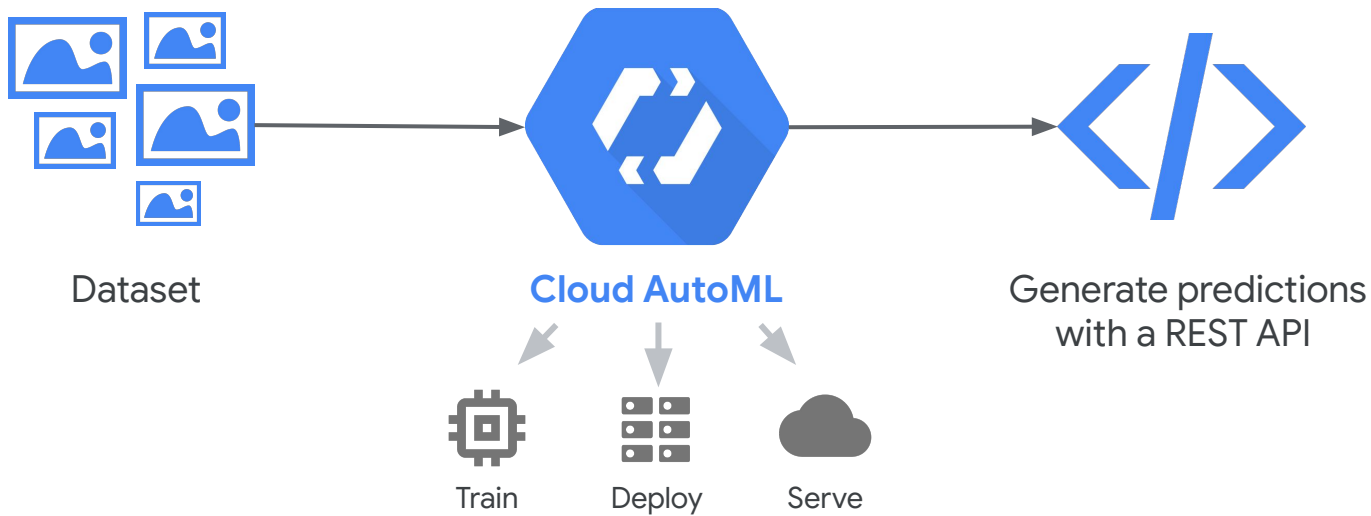






Large computational resources

Machine learning expertise



AutoML Vision

Upload and label images

Train your model

Evaluate



Handbag



Shoe



Hat



Cloud AutoML



Google Cloud Platform

AutoML Demo

Search products and resources

Vision

Dashboard

Datasets

Models

←

x_ray_pneumonia

LABEL STATS

EXPORT DATA

IMPORT

IMAGES

TRAIN

EVALUATE

TEST & USE

Single-Label Class

This dataset was copied to AI Platform (Unified) on Apr 6, 2021, 12:41:25 PM. Changes made to this resource won't affect the migrated copy. [Learn More](#)

GO TO UNIFIED AI PLA

All images

5,808

Labeled

5,808

Unlabeled

0

Filter

Filter label s

+

NORMAL

1,571

PNEUMONIA

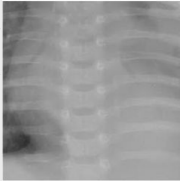
4,237

ADD NEW LABEL


Filter

Filter images

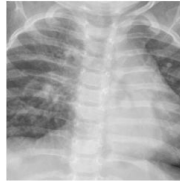
Select all




PNEUMONIA(1)




PNEUMONIA(1)




PNEUMONIA(1)




PNEUMONIA(1)




PNEUMONIA(1)




PNEUMONIA(1)




NORMAL(1)



PNEUMONIA(1)



NORMAL(1)



PNEUMONIA(1)

AutoML Natural Language

Upload and label text

Train your model

Evaluate



Sports



Lifestyle,
Money



Tech



Cloud AutoML



Sports

Lifestyle

Money

Tech

Travel

AutoML Tables

Upload labeled
structured data

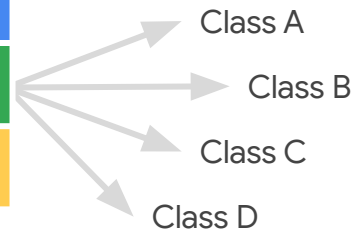
ID	Feature	Feature	...	Label
			...	
			...	

Train your model



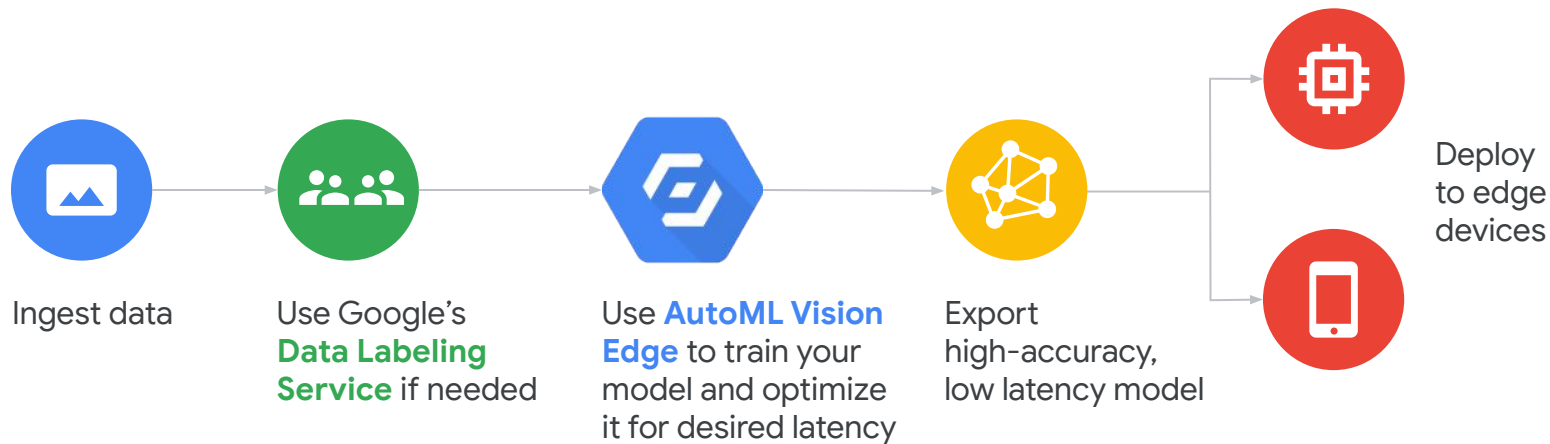
Evaluate (Classification
/ Regression)

ID	Feature	Feature	...
			...
			...



Model is now trained and ready to make predictions
This model can scale as needed to adapt to customer demands

AutoML Edge



AutoML Edge

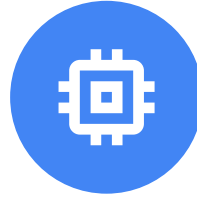
Easily build and deploy high-accuracy, low-latency models to edge devices



Classify images on-device and trigger actions in real-time, even with limited or unreliable connectivity



Achieve fast, state-of-the-art prediction accuracy for a range of applications



Supports a variety of edge devices



Ensure privacy by keeping data on the device, never leaving your enterprise premise

Demo time!



Comprehensive Suite for AI Developers

Sight



Cloud Vision



Cloud Video Intelligence



AutoML Vision



AutoML Video Intelligence

Language



Cloud Translation



Cloud Natural Language



AutoML Translation



AutoML Natural Language

Conversation



Dialogflow Enterprise Edition



Cloud Text-to-Speech



Cloud Speech-to-Text

Structured Data



AutoML Tables



Recommendation AI
(now GA!)



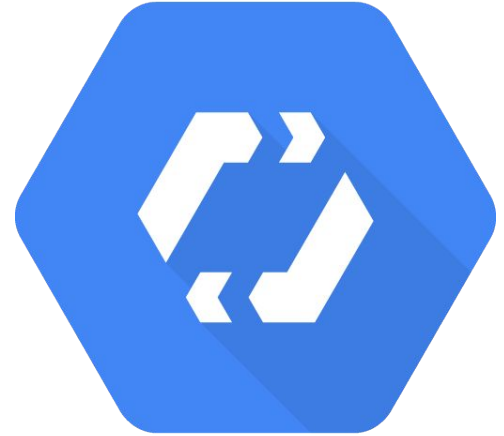
Cloud Inference API
(alpha)

Summary

Cloud AutoML offers **higher model accuracy and faster time to create** a production-ready model.

Users need **limited ML expertise to build a custom ML model**.

Simple graphical user interface to upload their images, train and manage their models, and then use those models directly on Cloud.



Customer asks...

Can we identify relevant objects from an image?

Vision API / AutoML Vision



Customer asks...

Can we predict the price of a house based on different attributes?

AutoML Tables
(regression)



Customer asks...

Can we personalize recommendations for our theme park visitors based on demographics, weather, etc ?

Recommendations
AI



Customer asks...

Can we extract patient data (name, insurance, etc)
from a form automatically?

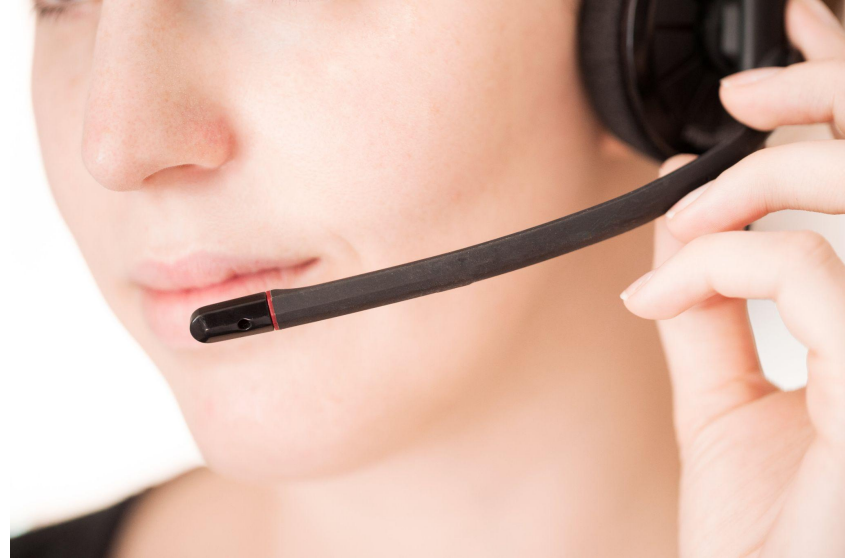
Document AI /
AutoML Vision



Customer asks...

Can we monitor financial product selling phone calls to identify if our agents are elaborating on our key products?

Speech to Text /
Text Classification



Customer asks...

Voice ordering assistant for our fast food mobile app channel?

Dialogflow



Today's demos

Vision API using the API Explorer

API Explorer: <https://developers.google.com/apis-explorer>

Vision API method images.annotate: <https://cloud.google.com/vision/docs/reference/rest/v1/images/annotate?apix=true>

Vision API feature types: <https://cloud.google.com/vision/docs/reference/rest/v1/Feature#type>

Natural Language API Python demo

Natural Language API client libraries: <https://cloud.google.com/natural-language/docs/reference/libraries>

Natural Language API code samples: <https://cloud.google.com/natural-language/docs/samples>

AutoML Vision demos

Cloud-hosted model quickstart: <https://cloud.google.com/vision/automl/docs/quickstart>

Edge device model quickstart: <https://cloud.google.com/vision/automl/docs/edge-quickstart>

Export edge model: <https://cloud.google.com/vision/automl/docs/export-edge> (TF Lite + android, [Edge Container](#))

Quick API demos

Document AI

<https://cloud.google.com/document-ai/docs/drag-and-drop>

Translate

<https://cloud.google.com/translate>

Vision AI

<https://cloud.google.com/vision>

Natural Language

<https://cloud.google.com/natural-language>

Speech to Text

<https://cloud.google.com/speech-to-text>

API & AutoML qwiklabs

[Image Processing](#)

[Analyzing a Tennis Serve with the Video Intelligence API](#)

[Building a Document AI Pipeline with Google Cloud](#)

[Integrate with Machine Learning APIs](#)

[Classify Images of Clouds in the Cloud with AutoML Vision](#)



BigQuery ML



Google BigQuery

Google Cloud's **enterprise data warehouse** for analytics

Exabyte-scale storage and
Petabyte-scale SQL queries

Encrypted, durable,
and highly available

Unique

Fully managed and **serverless**
for maximum agility and scale

Unique

Real-time insights
from streaming data

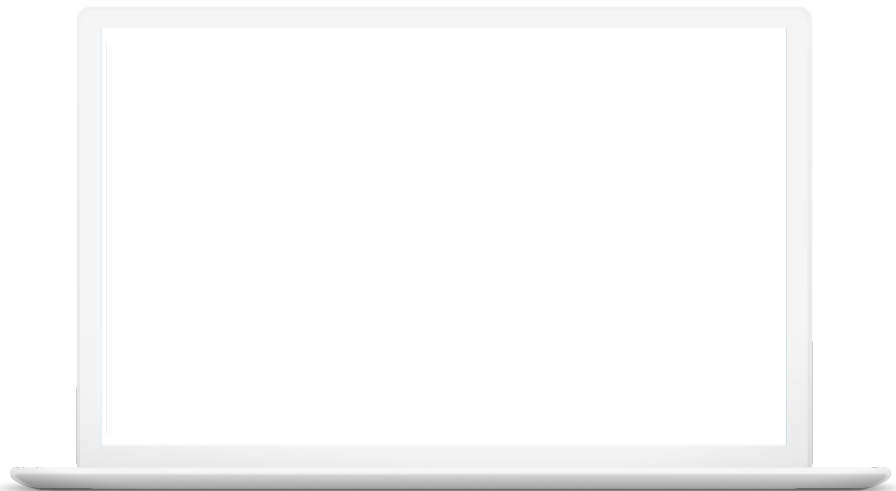
Unique

Built-in **ML** for out-of-the-box
predictive insights

Unique

High-speed, in-memory **BI Engine**
for faster reporting and analysis

BigQuery ML



Execute ML initiatives without moving data from BigQuery

Iterate on models in SQL in BigQuery to increase development speed

Automate common ML tasks

In-built infra management, security & compliance

Supported models

Classification

- Logistic regression
- DNN classifier (TensorFlow)
- Boosted trees using XGBoost
- AutoML Tables

Other Models

- k-means clustering
- Time series forecasting
- Recommendation:
Matrix factorization

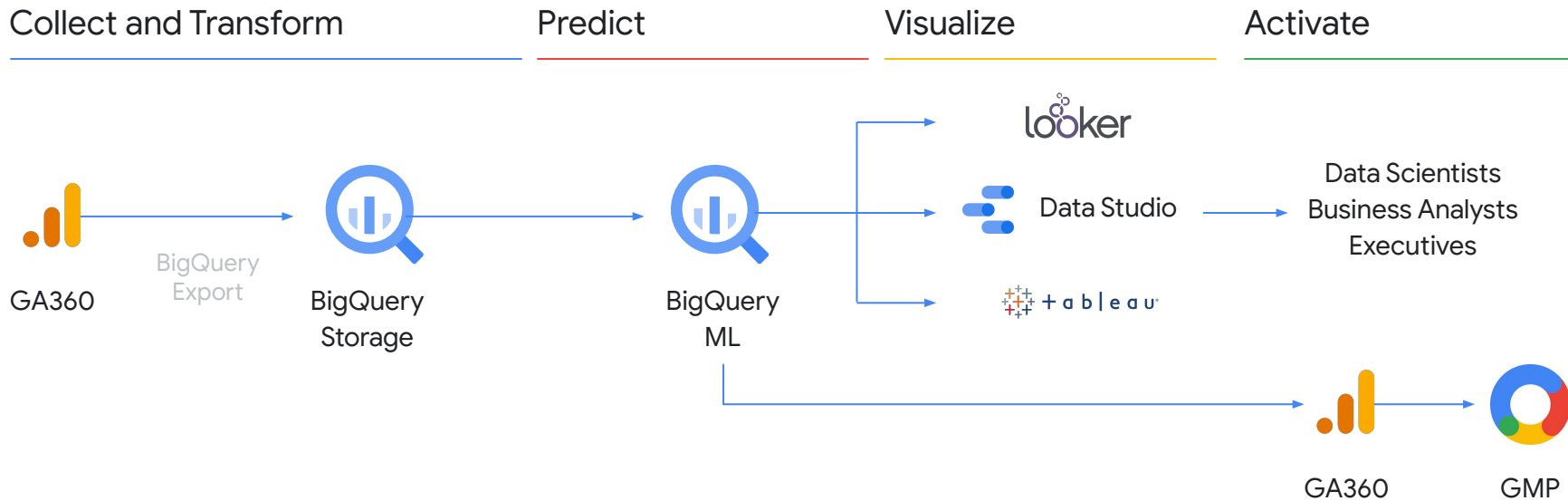
Regression

- Linear regression
- DNN regressor (TensorFlow)
- Boosted trees using XGBoost
- AutoML Tables

Model Import/Export

- TensorFlow models for
batch and online
prediction

Architecture



SQL interface to Machine Learning

- Data scientists can:
 - Build 100 models in a week
 - Not spend time on data ETL and setup of multiple tools
 - Experiment in BigQuery, export models to CAIP for further tuning
- Data analysts can:
 - Build ML models with knowledge of basic ML concepts.
 - No need to use multiple languages, just SQL
 - Easily utilize domain and data knowledge for ML models
- All: model can be served immediately after it's trained

AutoML tables through BigQuery ML

```
CREATE MODEL
```

```
  `demo.return_buyer_model`
```

```
OPTIONS
```

```
  (model_type='automl_classifier',  
   labels = ['will_buy_later'])
```

```
AS SELECT ...
```

```
SELECT
```

```
  fullVisitorId,  
  predicted_will_buy_later,  
  predicted_will_buy_later_probs
```

```
FROM ML.PREDICT(MODEL `demo.return_buyer_model`,  
  (SELECT ...
```

Export to GCS

```
PROJECT=$(gcloud config get-value project)
BUCKET=${PROJECT}-eu
gsutil mb -l eu gs://${BUCKET}
bq extract -m ch09eu.bicycle_model_linear \
           gs://${BUCKET}/bqml_model_export/bicycle_model_linear
```

Deploy to Cloud AI Predictions

```
./deploy.sh gs://${BUCKET}/bqml_model_export/bicycle_model_linear \
            europe-west1 london_bicycles bqml
```

Online Prediction

Run Online Predictions

```
gcloud ai-platform predict --model london_bicycles \
                           --version bqml --json-instances input.json
```

Demo time!



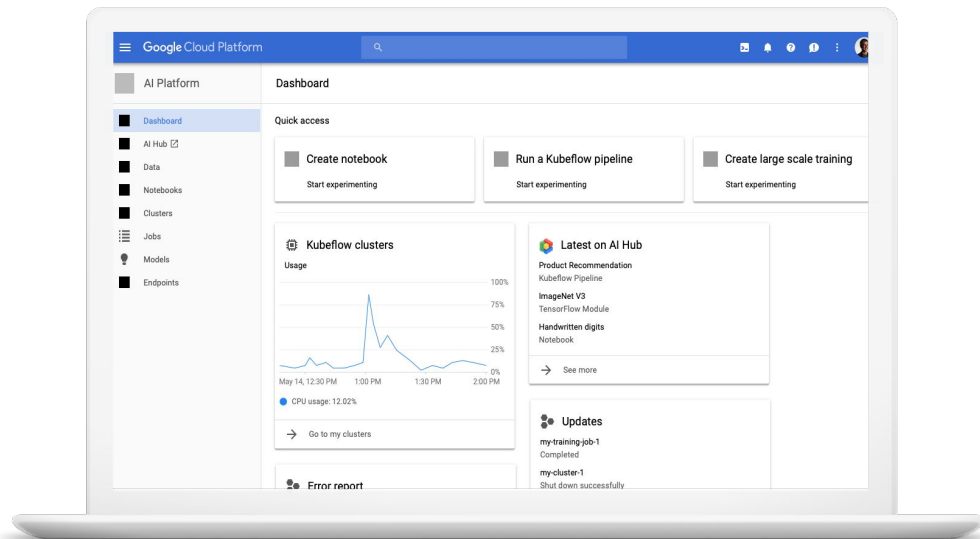


AI Platform

AI Platform (unified)

Proprietary + Confidential

- **End-to-end**, code-based development environment for AI inside GCP console built using expertise from Google's internal AI development
- **Individual AutoML** products will be integrated as features into Cloud AI Platform to create a cohesive workbench for ML builders
- Enterprise customers will benefit from a seamless, cutting-edge ML development experience from beginner to expert and for any ML use case



Our AI platform enables customers to build AI applications for their own organization leveraging Google's technology, scale and “way of doing things”.



AI Platform

Applications

Vision and Video

Conversation

Language

Structured Data

Core

Notebooks

Data Labeling

Deep Learning Env

Experiments

ML Metadata

AutoML

Training

Explainable AI

Feature Store

Vizier (Optimization)

Prediction

Model Monitoring

ML Pipelines

AI Accelerators

Hybrid AI

A unified object framework across AI Platform

Proprietary + Confidential

Dataset

- Unstructured or structured datasets housed within GCP with managed metadata including annotations
- Can be stored in GCS object storage, CSVs or BQ Tables

Training Pipeline

- Object created to train a model using a dataset
- Can also use external datasets in compatible formats

Model

- Machine learning model artifact with metadata built with a Training Pipeline
- Compatible models can also be uploaded directly to the platform

Endpoint

- Endpoint that can be called by users for online predictions and explanations
- Contains one or more deployed models which may exist independently from other model artifacts

AI Platform Notebooks

Proprietary + Confidential

Notebook instances

NEW INSTANCE

REFRESH

START

STOP

RESET

DELETE

Filter table

Instance name

Customize instance

R 3.6

R 3.6 and key libraries pre-installed

Python

Python 2 and 3 with Pandas, SciKit Learn and other key packages pre-installed

TensorFlow Enterprise 1.15

TensorFlow Enterprise 1.15 pre-installed with support for Keras

TensorFlow Enterprise 2.1

TensorFlow 2.1 pre-installed with support for Keras

PyTorch 1.4

PyTorch 1.4 pre-installed

RAPIDS XGboost [EXPERIMENTAL]

XGboost optimized for NVIDIA GPUs

CUDA 10.1

Optimized for NVIDIA GPUs

Without GPUs

With 1 NVIDIA Tesla K80

- Schedule execution of notebooks
- **Integration with GCP:** BigQuery and GCS data exploration, execution with CAIP training
- Ability to **scale hardware** up or down directly from JupyterLab.

- **Secure development:** private networking, VPC-SC, shared VPC and private IP controls. You can also encrypt your data on disk with CMEK.
- **Controlled user access:** single-user or service account
- **Support for data science frameworks**, including R, pandas, NumPy, SciPy, scikit-learn, and Matplotlib, and ML frameworks like TensorFlow, Keras, fast.ai, RAPIDS, XGBoost, and PyTorch.

Download
BigQuery data to
our notebook

```
query="""  
SELECT  
    weight_pounds,  
    is_male,  
    mother_age,  
    plurality,  
    gestation_weeks  
FROM  
    publicdata.samples.natality  
WHERE year > 2000  
LIMIT 10000  
"""  
  
df =  
bigquery.Client().query(query).to_dataframe()  
df.head()
```

Demo time!



AI Platform prices - AutoML

Proprietary + Confidential

For AutoML models, you pay for three main activities:

- Training the model
- Deploying the model to an endpoint
- Using the model to make predictions

Image data

Video data

Tabular data

Text data

Operation	Price per node hour (classification)	Price per node hour (object detection)
Training	\$3.15	\$3.15
Training (on-device)	\$4.95	\$18.00
Deployment and online prediction	\$1.25	\$1.82
Batch prediction	\$2.02	\$2.02

AI Platform prices - Training and Prediction

Proprietary + Confidential

Price per node hour, depending on type machine, accelerator and disk type.

Example for prediction:

Americas <u>Europe</u> Asia Pacific	
Prediction machine type SKUs	
vCPU	\$0.036632 per vCPU hour
RAM	\$0.004902 per GB hour

You can optionally use [GPU accelerators for prediction](#). GPUs incur an additional charge, separate from those described in the previous table. The following table describes the pricing for each type of GPU:

Americas <u>Europe</u> Asia Pacific	
Accelerators - price per hour	
NVIDIA_TESLA_K80	Not available
NVIDIA_TESLA_P4	\$0.6500
NVIDIA_TESLA_P100	Not available
NVIDIA_TESLA_T4	\$0.3800
NVIDIA_TESLA_V100	\$2.5500

AI Platform prices - Data Labeling

Proprietary + Confidential

Price per 1,000 units per human labeler (Tier-1, tier-2 are volumes)

Data type	Objective	Unit	Tier 1	Tier 2
Image	Classification	Image	\$35	\$25
	Bounding box	Bounding box	\$63	\$49
	Segmentation	Segment	\$870	\$850
	Rotated box	Bounding box	\$86	\$60
	Polygon/polyline	Polygon/Polyline	\$257	\$180
Video	Classification	5sec video	\$86	\$60
	Object detection	Bounding box	\$86	\$60
	Object tracking	Object in 30sec video	\$686	\$480
	Action recognition	Event in 30sec video	\$214	\$150
Text	Classification	50 words	\$129	\$90
	Sentiment	50 words	\$200	\$140
	Entity extraction	Entity	\$86	\$60
Active Learning	All	Data item	\$80	\$56
Custom Labeler	All	Data item	\$80	\$56

Resources

AI Platform Product Page



cloud.google.com/ai-platform

ML Engineer Certification



cloud.google.com/certification/machine-learning-engineer

TensorFlow Extended



tensorflow.org/tfx

KubeFlow Pipelines



kubeflow.org/docs/pipelines/

Gracias