

Hands on AlphaFold with Vertex AI

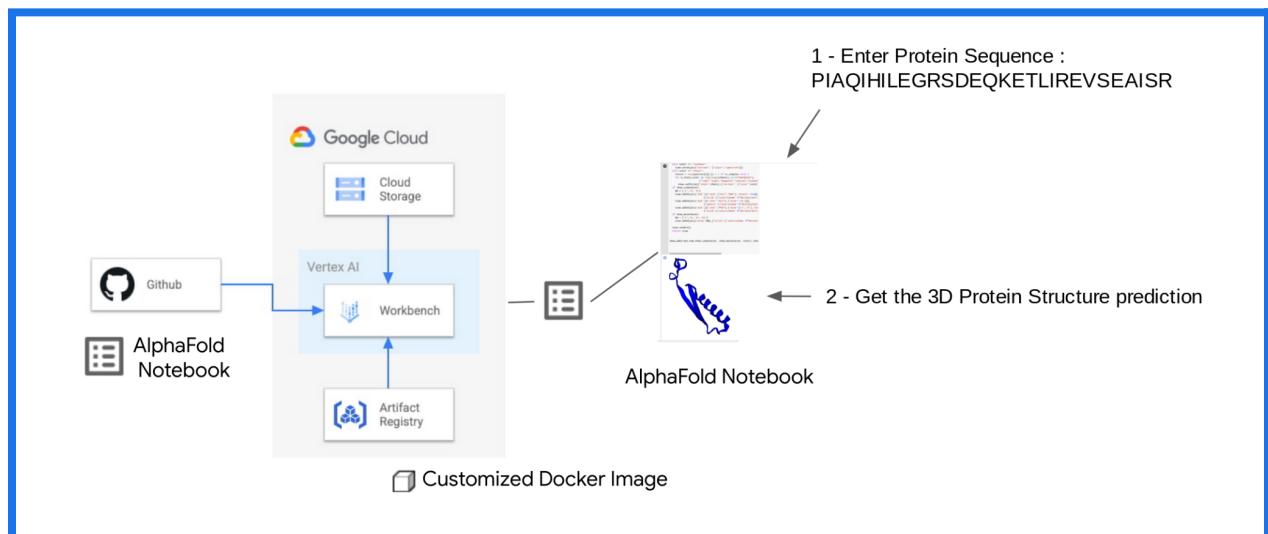
4th August 2022

AlphaFold is Deepmind's Machine Learning based protein structure prediction system.

Within this hands-on session we will walk through a demonstration notebook that can predict the structure of a protein using a slightly simplified version of AlphaFold that does not use homologous protein structures or the full-sized BFD sequence database.

Please note that this session is provided as an early-access prototype and is not reflecting the quality of a finished product. It is provided for theoretical modeling only and caution should be exercised in its use.

Hands On Overview:

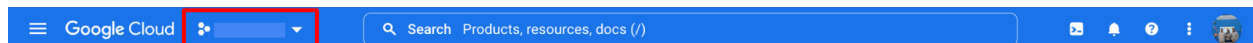


[1] Open a new incognito tab and open this guide in the incognito session

[2] Go to: <https://console.cloud.google.com/>

[3] Connect with the provided credentials

[3.1] Ensure you select the “project” by selecting it in the top blue bar drop down menu



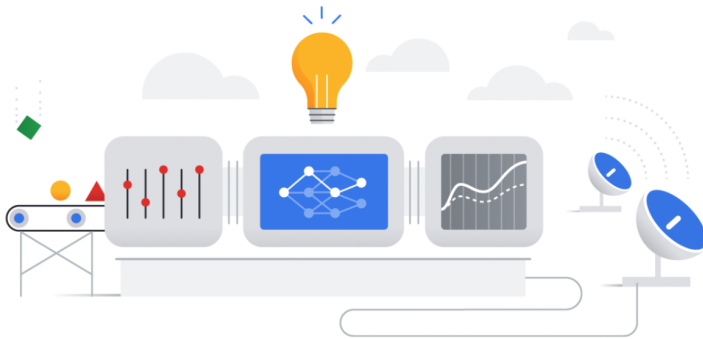
[4] Enable APIs first, in the top head search bar, search Vertex AI , then click on Vertex AI:

[5] Click the **Enable VERTEX AI API** button

Get started with Vertex AI


Vertex AI empowers machine learning developers, data scientists, and data engineers to take their projects from ideation to deployment, quickly and cost-effectively. [Learn more](#)

ENABLE VERTEX AI API




The illustration depicts a machine learning pipeline. It starts with a data input (a green diamond and a yellow circle), followed by a model training stage (a box with red vertical bars), a prediction stage (a box with a neural network diagram), and a deployment stage (a box with a line graph). The pipeline is connected to a satellite dish, symbolizing remote access or deployment.

Region
us-central1 (Iowa) ▼ ⓘ

 **Prepare your training data**


Collect and prepare your data, then import it into a dataset to train a model

+ CREATE DATASET

 **Train your model**

Train a best-in-class machine learning model with your dataset. Use **Google's AutoML**, or bring your own code.

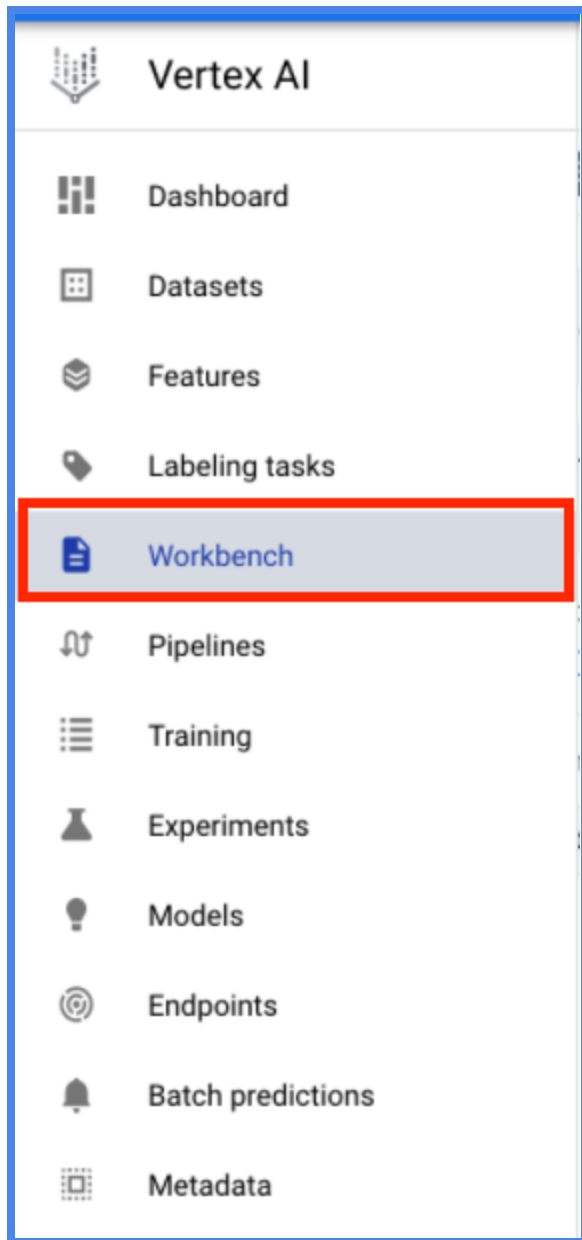
+ TRAIN NEW MODEL

 **Get predictions**

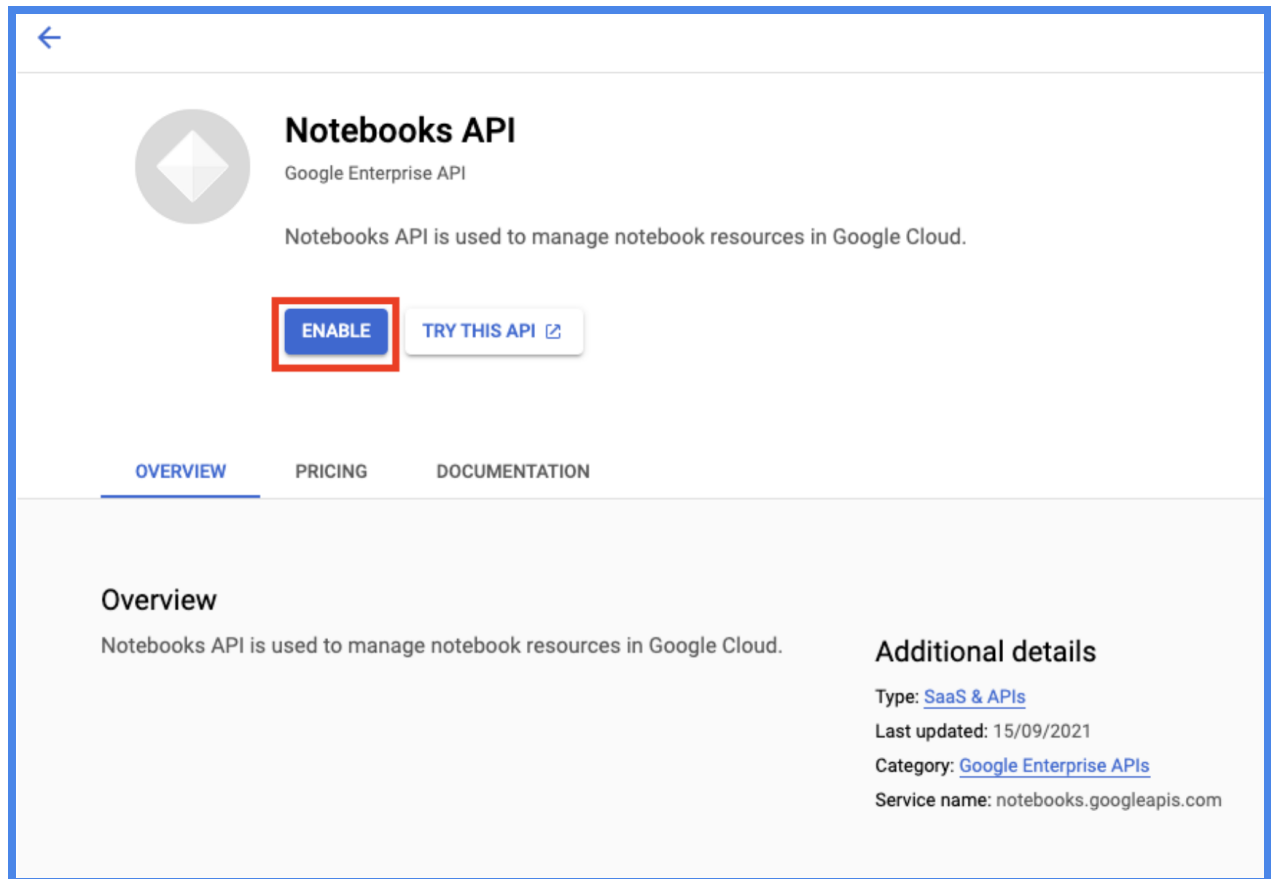
After you train a model, you can use it to get predictions, either online as an endpoint or through batch requests

+ CREATE BATCH PREDICTION

[6] On the left side Navigate in Workbench:



[7] Enable the NotebookAPI by clicking the **Enable** button



The screenshot shows the Google Cloud console page for the Notebooks API. At the top left is a back arrow. Below it is a circular icon with a diamond shape. To the right of the icon is the title "Notebooks API" and the subtitle "Google Enterprise API". A description states: "Notebooks API is used to manage notebook resources in Google Cloud." Below the description are two buttons: "ENABLE" (highlighted with a red box) and "TRY THIS API" with an external link icon. A navigation bar contains three tabs: "OVERVIEW" (selected), "PRICING", and "DOCUMENTATION". The "Overview" section contains the same description. To the right, under "Additional details", are the following fields: "Type: [SaaS & APIs](#)", "Last updated: 15/09/2021", "Category: [Google Enterprise APIs](#)", and "Service name: notebooks.googleapis.com".

[8] Get the Notebook by visiting :


https://github.com/GoogleCloudPlatform/vertex-ai-samples/blob/main/community-content/alphafold_on_workbench/AlphaFold.ipynb

723 Lines (723 sloc) | 38 KB

<> Raw Blame

AlphaFold On Vertex AI Workbench

Vertex AI Workbench offers an end-to-end notebook-based production environment that can be preconfigured with the runtime dependencies necessary to run AlphaFold on Vertex AI. With [User-Managed Notebooks](#), you can configure a GPU accelerator to run AlphaFold using Tensorflow, without having to install and manage drivers or JupyterLab instances. This notebook allows you to easily predict the structure of a protein using a slightly simplified version of [AlphaFold v2.1.0](#).

 [Launch this Notebook in Vertex AI Workbench](#)

Differences to AlphaFold v2.1.0

In comparison to AlphaFold v2.1.0, this notebook uses **no templates (homologous structures)** and a selected portion of the [BFD database](#). We have validated these changes on several thousand recent PDB structures. While accuracy will be near-identical to the full AlphaFold system on many targets, a small fraction have a large drop in accuracy due to the smaller MSA and lack of templates. For best reliability, we recommend instead using the [full open source AlphaFold](#), or the [AlphaFold Protein Structure Database](#).

This notebook has a small drop in average accuracy for multimers compared to local AlphaFold installation, for full multimer accuracy it is highly recommended to run AlphaFold locally. Moreover, the AlphaFold-Multimer requires searching for MSA for every unique sequence in the complex, hence it is substantially slower. If your notebook times-out due to slow multimer MSA search, we recommend running AlphaFold locally.

Please note that this notebook is provided as an early-access prototype and is not a finished product. It is provided for theoretical modelling only and caution should be exercised in its use.

[9] Continue the Notebook creation Selecting a Name typing it in the **Notebook name** section, select **Europe West 1** as Region and **don't click "Create" just yet** :

Google Cloud Platform alphafold-on-gcp Search products and resources

Vertex AI Deploy to notebook

To deploy this file, create a notebook with the required environment.

Notebook name *
alphafold-on-gcp-notebook

63 char limit with lowercase letters, digits, or '-' only. Must start with a letter. Cannot end with a '-'.

Region *
europe-west1 (Belgium)

Zone *
europe-west1-c

Notebook properties

Environment	TensorFlow Enterprise 2.6 (with LTS and Intel® MKL-DNN/MKL)
Machine type	4 vCPUs, 15 GB RAM
Boot disk	100 GB Standard persistent disk
Data disk	100 GB Standard persistent disk
Subnetwork	default(10.138.0.0/20)
External IP	Ephemeral(Automatic)
Permission	Compute Engine default service account
Estimated cost	\$102.70 monthly, \$0.141 hourly

ADVANCED OPTIONS CANCEL CREATE

[10] Click **ADVANCED OPTIONS** section

[11] Select **Custom container** in the **Environment** drop down from the **Environment** section

[12] In the “Docker container image” text box, enter (without clicking “select”):

`us-west1-docker.pkg.dev/cloud-devrel-public-resources/alphafold/alphafold-on-gcp:latest`

[13] Machine type drop down : Machine type: n1-standard-8 (8 CPUs, 30 GB RAM)

[14] GPU Type drop down : NVIDIA Tesla T4

[15] check box “Install NVIDIA GPU driver automatically for me”

Google Cloud

Search Products, resources, docs (/)

Vertex AI

Deploy to notebook

Environment

All environments have the latest NVIDIA GPU libraries, Intel libraries and drivers. Notebooks use JupyterLab 3 by default (you can [specify a previous version](#) instead).

Environment *
Custom container

Supply your own custom container environment. The container should conform to the requirements listed [here](#).

Docker container image *
us-west1-docker.pkg.dev/cloud-devrel-public-resources/alphafold/alphafold SELECT

Health check path

Health check interval
0 seconds

Select a script to run after creation BROWSE

Custom metadata ?

Metadata

+ ADD ITEM

Machine configuration

Machine type *
n1-standard-8 (8 vCPUs, 30 GB RAM)

GPU type
NVIDIA Tesla T4

Number of GPUs
1

☒ Install NVIDIA GPU driver automatically for me ?

[16] Press **Create** !

[17] After several minutes, a virtual machine will be created and you will be redirected to a JupyterLab instance. When launching, you may need to confirm the connection to the Jupyter server running on the VM; click the **Confirm** button

Confirm deployment to notebook server

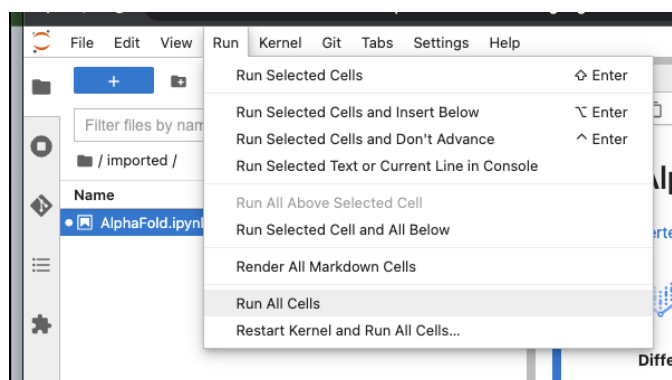
Are you sure you want to download the notebook source from:

https://github.com/GoogleCloudPlatform/vertex-ai-samples/raw/alphafold_on_gcp/community-content/alphafold_on_workbench/AlphaFold.ipynb

Confirm

[Return to JupyterLab](#)

[18] The notebook is ready to run! From the menu, select **Run > Run all Cells** to evaluate the notebook top-to-bottom, or run each cell by individually highlighting and clicking **<shift>-return**. The notebook has detailed instructions for every step, such as where to add the sequence(s) of a protein you want to fold.



[19] Congratulations, you've just folded a protein using AlphaFold on the Vertex AI Workbench!

