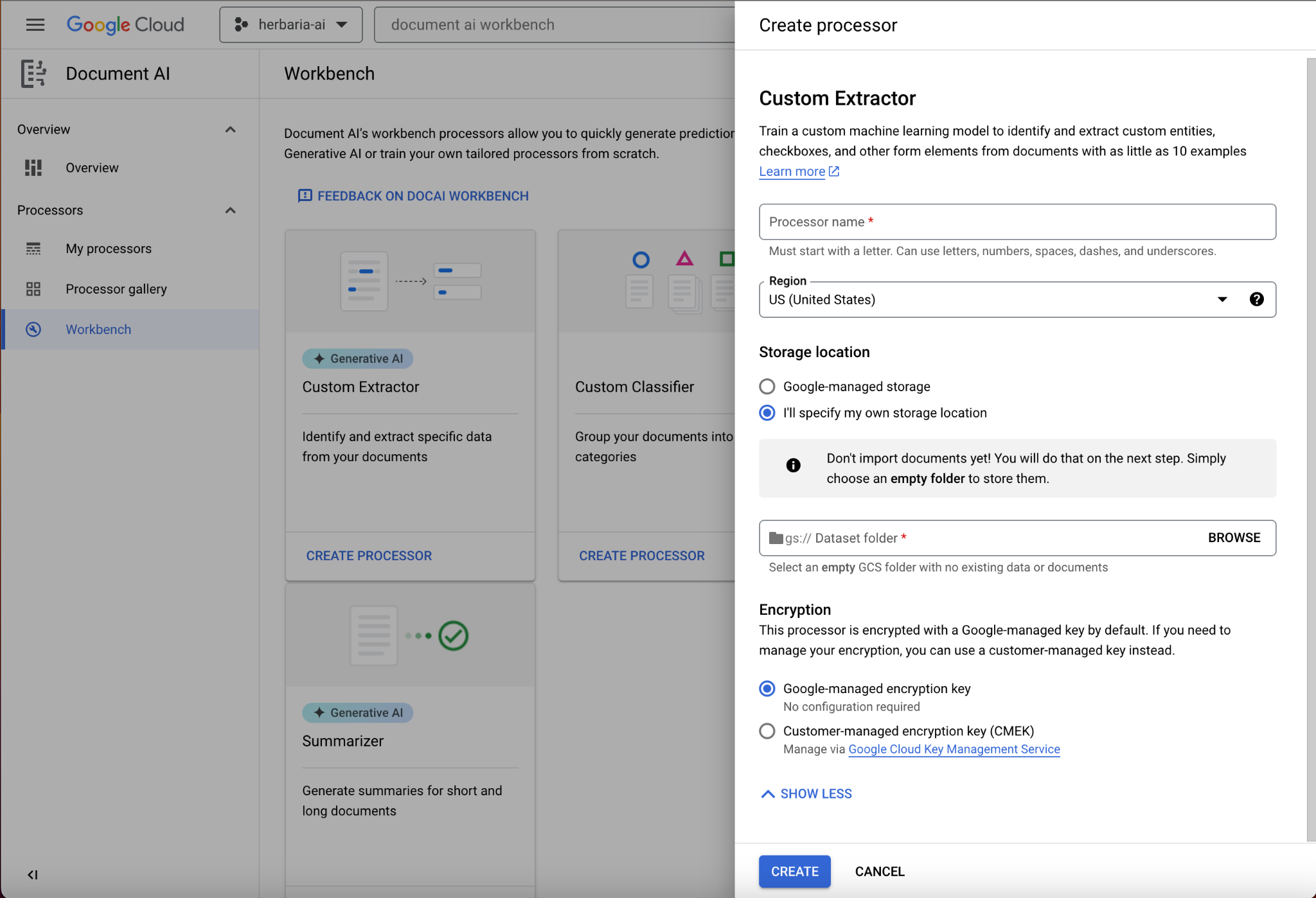
# Create bucket and upload files

# Open Document AI Workbench and Create a Processor



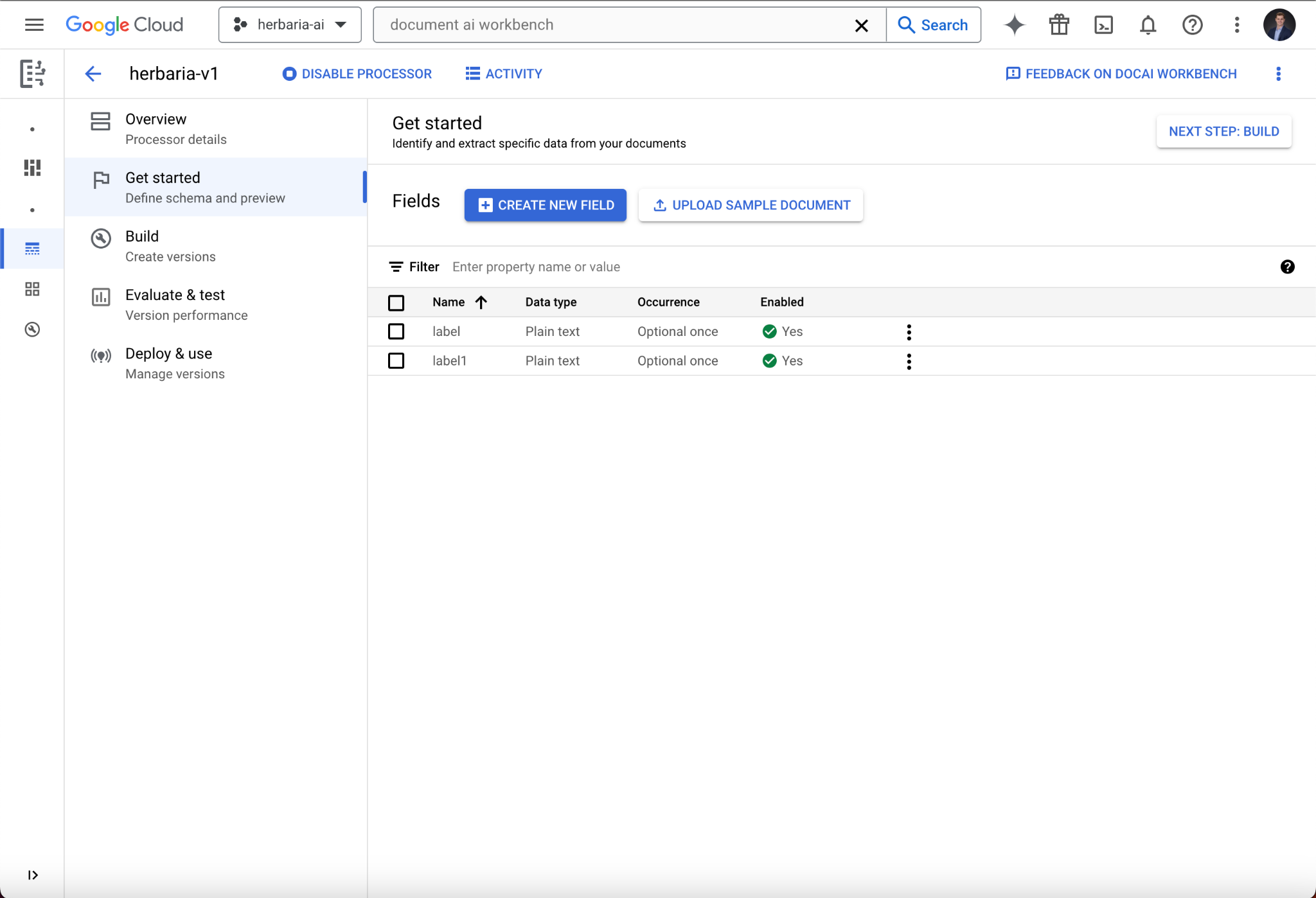
Specify the Storage Location you would like to use for results (must be an empty folder)



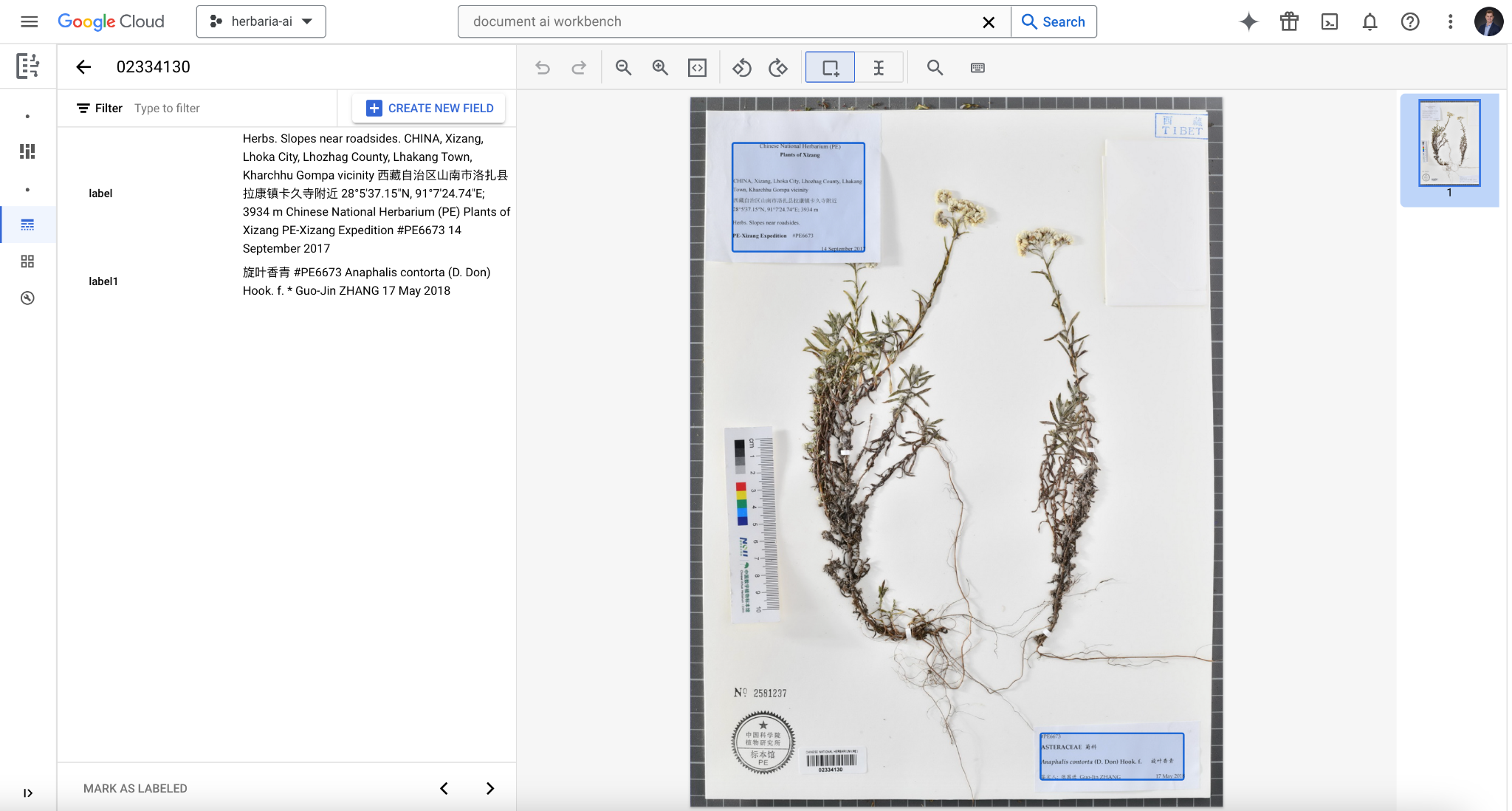




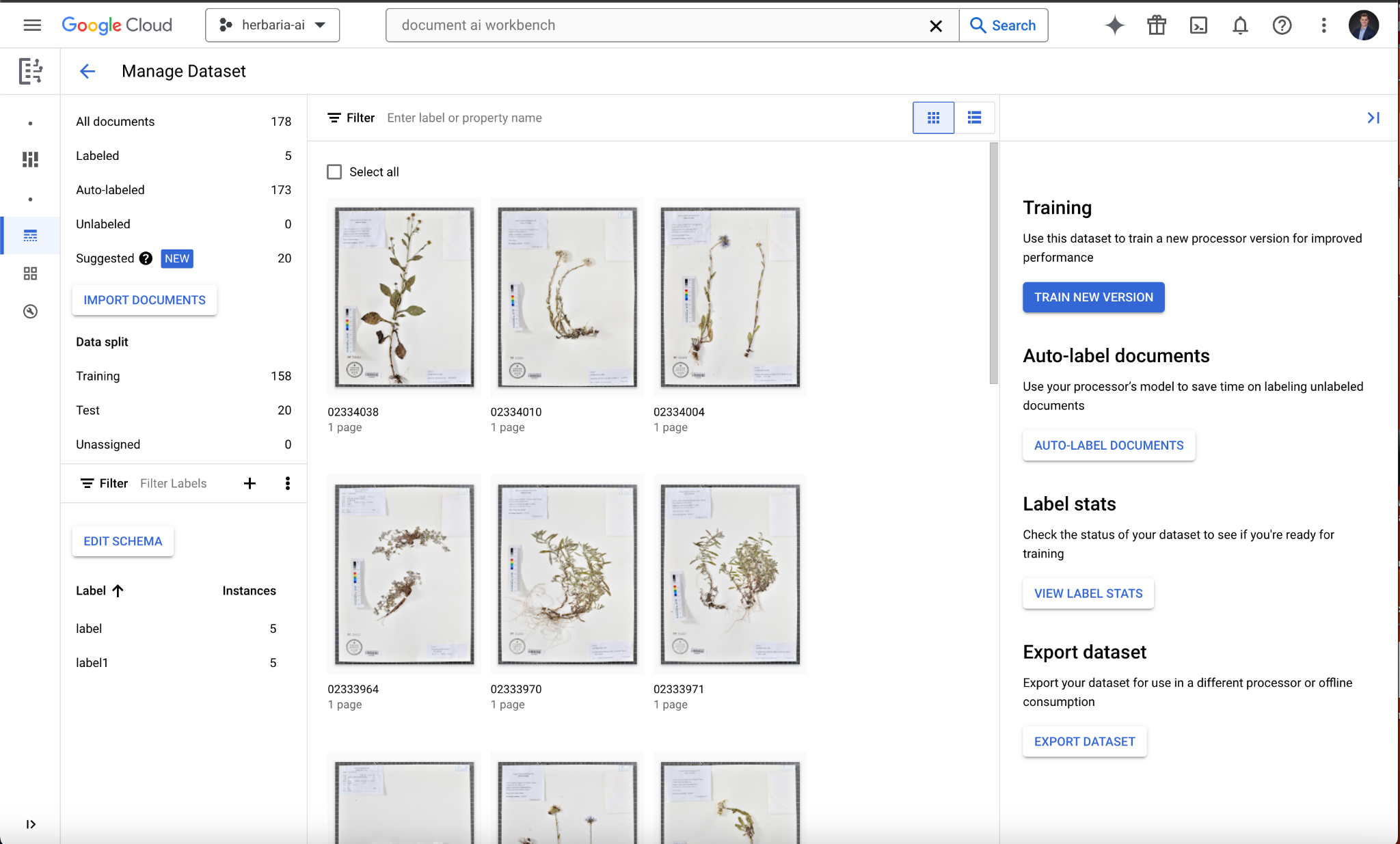
# Create Labels and Upload Training/Test Documents

Labels can be named and you can choose the number of possible occurrences they can have. Google uses Generative AI to place bounding boxes on the document (see example below) based on a sample document you can upload. We will further explore how detailed we can be with these labels and how flexible they are to different document layouts

# 



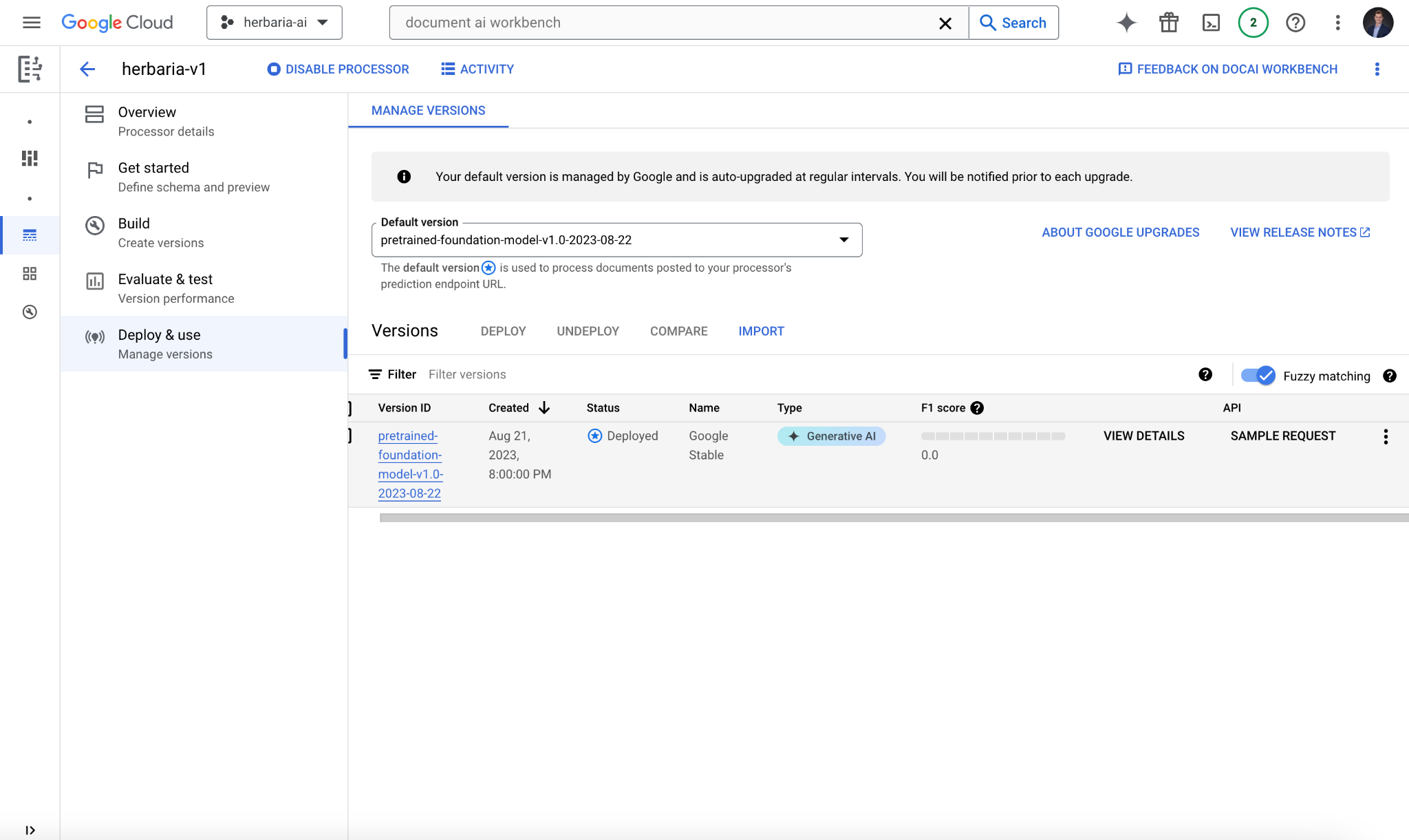
# Import train/test documents and label

You can then import documents to train the model and test your labeling. You can either specify a standard train test split or upload all documents as training data and later move a specific number of them to test data. You can either auto-label the documents you upload based on the 

sample labeled document you created or you can label by hand certain documents.



# Deployment

Once your model is trained and evaluated it can then be deployed using the processor's prediction endpoint URL. The model can of course be retrained and reevaluated at any time as new documents are added or as labels are added or modified.

# Next Steps

We will continue finding ways to streamline this pipeline and improve the results of the labeling and training process. We have also enabled the Google Translation API and will look into implementations of it within Document AI itself. As well as test calling our processor in a Python notebook.