https://docs.google.com/document/d/1UdvE4wuVWBTIKL8oMEoWnCWqMy6FM7y-LE6dCsH2iGY/edit#

[Google Confidential - Under NDA] - April 1, 2021

**Table of Contents**

[**Overview and Concepts**](#_ehgsc9ip1tcm) **2**

[Product Overview](#_avmtct5u6sok) 2

[Features](#_si1w2b2tz8x9) 2

[Benefits](#_8q93dfu32z2n) 2

[Processors supported](#_fqoarc3fmr6t) 3

[Availability](#_tw7aclrca98x) 3

[Concepts](#_lb5n9x69bs1g) 3

[Best Practices and Known Limitations](#_w5pj94nwx3a7) **4**

[Best Practices](#_rhei6gub9x0v) 5

[Known Limitations and Workarounds](#_xr1nwkd2isy5) 6

[**Quickstart: Configuring a HITL Document Review Task**](#_h19jo9qhfwfn) **7**

[Customer Steps](#_utse8dg7qjc5) 8

[Step 0: Enable the API Platform (Unified) API](#_o1y2tdvgxzvk) 8

[Step 1: Create a processor and set up human review](#_poc6ngiw4auc) 8

[Step 2. Configure Human Review](#_x0orfl69ibyh) 10

[Labeling Manager Steps](#_howld9tdvfjm) 13

[Step 3: Add Labelers](#_1v5tq1hmm2ww) 13

[Step 4: Post an Initial Document to Review](#_26tlgt4swyxs) 15

[Step 5: Review the document](#_209fetz7fr2u) 16

[Labeller Steps](#_2qvv5stamdrc) **16**

[Step 6: Labeller Procedure](#_zgvmp0t6lbdr) 16

[Step 6: [Manager] Review Document from Cloud Storage](#_175w57xowa7d) 22

[Step 7: [Manager] View HITL Review Analytics by Task and by Labeler](#_26tlgt4swyxs) 22

[Labeling Manager: Manage your labeling tasks in DataCompute](#_tanyh12bk5pi) **23**

[Manage Human Review Queues](#_3pn88v2ticig) 23

[Receive Email Notifications](#_dt6fx7vkbpa) 25

[Add Labelers](#_kl3ff9m0n2q2) 25

[Assign Tasks](#_z31q9a9ft4iw) 26

[Notify Labelers](#_m7gf03tisard) 26

[**Google HITL Workforce [Private Preview]**](#_uc7yyk9phjca) **28**

[HITL Review Center Features](#_r93v619iy0qw) 28

[Information security](#_c89lj0xh0ogf) 28

[Personnel security](#_4o2vnk3othr0) 28

[Physical security](#_2bt672sk9iq) 28

[COVID-19 Imposed Data Restrictions](#_tvmd981zguj) 29

[Currently the Labeling centers are not open and the HITL workforce is working from home. As a result, they’re restricted from reviewing any documents containing PII. If your documents contain PII fields, we are restricted from offering this service until the labelers start working from the secure labeling center facilities.](#_dlxpk6e0k2uk) 29

[Availability](#_sgtcicup3vg3) 29

[**HITL Pricing**](#_orlxp3bi0wu3) **29**

[Platform Pricing](#_5klcs1ipahpt) 29

[Workforce Pricing](#_44nyiwpl03yn) 29

[**Labeler Workbench**](#_h8k75c4pdwem) **29**

[**Labeling Manager Console**](#_o0q0kyw86xrs) **30**

[**Creating Instructions for HITL Review**](#_4jq0zbrfttm7) **30**

[Design good instructions](#_bzel6adww2bn) 30

[Visual Examples](#_cwic2rxxgafx) 31

# Overview and Concepts

[LINK - [https://cloud.google.com/document-ai/docs/hitl/concepts]](https://cloud.google.com/document-ai/docs/hitl)

## Product Overview

HITL AI enables human verification and corrections to ensure accuracy of data extracted by DocAI processors, before it is used in critical business applications. It provides a workflow and UI for humans (referred to as labelers in HITL) to review, validate and correct the data extracted from documents by DocAI processors. We provide 2 options of human labelers

* Bring-your-own-labelers - your own workforce or partner workforce can review the documents
* Google HITL Workforce - is available as a Private Preview service for documents. Currently, this is limited to review of documents that contain no PII ([sign up form](https://docs.google.com/forms/d/1l-utSiZJ51evNo6XF5li-4QR5tcmfEeRTgpIRYRUcys/viewform?edit_requested=true)).

### Features

We support the following features

* Confidence threshold filters to limit the number of documents going through HITL
* Labeler pool management, including task assignments to optimize load balancing, escalations, SLO priorities
* Efficiency analytics by task and by labeler
* UI cues and features that reduce labeler handling time per document
* Analytics and metrics by task and by labeler, so you can streamline HITL operations

### Benefits

HITL AI offers the following benefits

* **Risk mitigation** - Could mitigate risks of critical data being incorrect - e.g. invoice amounts, billing addresses, loan amounts etc.
* **Simplify exception handling** - Easily roll out a human review and exception handling workflow.
* **Workforce efficiencies** - Manage, monitor and improve productivity of workforce managing human review.
* **Cost control** - Control costs of human review with configurable filters and
* **Data completeness** - Ensure extracted data is complete for downstream business applications

### Processors supported

Currently, we support the HITL Review workflow for the following processors, and you can see the HITL configuration menu for such processors:

* Invoices
* Receipts

HITL for Lending AI and other parsers will be added in the product shortly - look out for the HITL Configuration menu for these.

### Availability

HITL Platform is in Private GA. Please fill out this [request form](https://docs.google.com/forms/d/1IaJh381nhXAuE0Ywc0msEaZKQgS7P5j0kciERJT8JOM/edit) if you would like to access this service.

## Concepts

Following are some concepts and features used in this product:

| **Concept** | **Definition** |
| --- | --- |
| Review | The process of visually comparing the extracted field values against values in the document and correcting any incorrect extractions or adding missed extracted fields missed by the DocAI processors. |
| Labeler | Human that reviews the extracted document. The customer can use their own workforce [Bring-your-own-labeler or BYOL] or use Google labelers for HITL Review |
| Task | A queue of extracted documents that labelers review. A processor generates a single task, when configured for HITL Review. |
| [Labeler Workbench](https://cloud.google.com/document-ai/docs/hitl/labeler-workbench) | UI used by Labeler to review documents. The UI presents documents from the queue one after another, that the labeler can review, correct and either submit or reject.   * BYOL labelers need to have a Google Workforce or Gmail account to access the labeling UI * Labelers can access the Workbench through a link sent via email by Labeling Manager upon task assignment, |
| Answer Time | This is the time taken by a labeler to process a document. The Labeler Workbench clocks each document submission and presents efficiency analytics (e.g. for each labeler document review) |
| Labeling Manager | One or more labeling managers are assigned to a pool of labelers, so that they can:   * Add/remove labelers to labeler pools and * Assign/unassign tasks to a labeler. All tasks in the project are accessible to a labeler manager. (S)he may change task assignments to labelers based on the changing priorities of tasks. * Pause tasks so that labelers can work on the next tasks assigned to them.   In the BYOL scenario, Labeling Managers are provided by the customer.  When Google labelers are used, Google provides the Labeling Manager. |
| [Labeling Manager Console](https://cloud.google.com/document-ai/docs/hitl/lm-console) | UI used by Labeling Manager to manage labeler pools and task assignments (as shown in screenshot below)   * The Console is accessible at <https://datacompute.google.com/cm> |
| Enqueue, Answered, Completed, Rejected Documents in a Task | A task is a continual workflow. A document goes through the following states:   * Enqueued - As documents are processed by the processor, they’re “enqueued” (added) to the HITL task * Answered - when a document is reviewed, corrected and submitted by a Labeler, it is “completed” and saved in the customer’s configured GCS * Completed - when a document is answered by all Labelers if the task has replication activated (i.e. multiple labelers working on each document in the task). When the task has no replication (i.e. reviewed by a single labeler), Answered is the same as Completed. * Rejected - a document may be rejected if it is an invalid document [different doc-type, forged, etc] or poor quality [glare, edge cut off, etc]. |
| Single Task per Processor | We do not support multiple tasks per processor. If customers need to process a single document type (say Invoices) in different tasks, they can configure multiple processors with HITL Review. |
| Task Assignment vs Labeler Pools | Labeling Manager adds labelers to a pool. Once added, any labelers from the pool can be assigned to a task.  Note, “Labeler pool” is not to be confused with the “group” of labelers assigned to a task. A Pool is managed at a Project level and is used to determine labeler access to the analytics and the tasks. Any labeler from the pool can be assigned to one of more tasks in the Project. |
| Labeler Pool | A pool of labelers is created at a project level and not to be confused with task assignments. The Labeling Manager can assign any Any labeler Any labeler assigned to a task, so that multiple labelers can review documents in parallel and complete the task quicker. A labeler pool can be assigned to any task in the project by the customer. |
| Validation filters and thresholds | Extracted fields have a confidence score [0-100] representing the confidence that the DocAI extraction is accurate. Customers can configure validation threshold for each field, so that only pages with fields that are below this validation threshold are enqueued for review, the rest are not.  There are 3 types of validation filters customers can configure:   * Field-level filter - select the important fields that need to be reviewed and specify a confidence threshold for each field. If this threshold is set at 100% for any field, all pages containing this field are sent for review. * Document-level filter - select an overall document-level confidence threshold. If any field is below the threshold, the entire page is sent for review. If this threshold is set at 100%, all documents predicted are sent for review. * No filter - every document posted to the HITL end-point is sent for review. |
| [Labeler Manager Analytics](https://cloud.google.com/document-ai/docs/hitl/lm-console) | The Labeling Manager gets analytics for each Task and each Labeler, including Enqueued, Answered, Skipped, Completed, Average Handling Time/document and total Answer time (as shown in screenshots below).  Analytics are accessed in the Analytics tab of Labeling Manager Console - <https://datacompute.google.com/cm> |

# **Best Practices and Known Limitations**

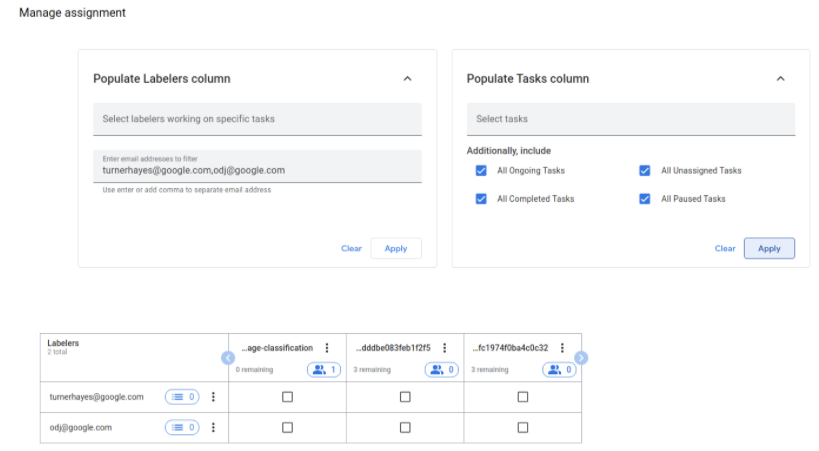
[LINK - [https://cloud.google.com/document-ai/docs/hitl/best-practices]](https://cloud.google.com/document-ai/docs/hitl)

## Best Practices

* **Single Labeler pool per project -** In the BYOL scenario, although the HITL Setup UI supports creating new labeler pools for each processor, it is recommended to use a single labeler pool across all processors in a project to keep things simple. This is because users from multiple labeler pools can still be assigned to the same task, and assigning a labeler pool to a task doesn’t restrict the Labeling Manager from assigning other labelers to that task.
* **Multiple tasks/document-type** - Multiple tasks may be required to process a single document-type. This is required in the following scenarios:
  + Labelers with different skills or credentials required to review/validate different fields of the same document
  + Documents with different fields (e.g. invoices from different vendors) may need different schema to be validated, hence need to be routed to different tasks
  + Documents from different customers have different labeler pools working on them (e.g. invoices from different customers, languages or countries)

In this scenario, we recommend creating multiple processors (and tasks) and posting the documents to the appropriate task. [A classifier may be required to classify these documents and post to the correct processor].

* **Limit Review to the fields required (vs all extracted fields on page), to save Review time and cost -** TheAnswer time/page scales with the number of fields to be reviewed. The Google labeler costs are based on the # of fields reviewed. Therefore it is recommended to limit the fields to be reviewed to the ones used in downstream business process.
  + For example, an Invoice might have 30+ fields but you may want to limit Review to only 4-5 fields that are important for settling the invoice.
* **Naming HITL-enabled Processors** - the HITL task name is the same as the processor name. It is recommended to use a name that is understood by the labeling managers and labelers, and easily distinguished from other tasks they would be working on.
* **Prioritizing Tasks** - if a labeler is assigned to multiple tasks, they process these sequentially. If a labeler needs to switch tasks (e.g. to respond to a customer escalation or meet an SLO), the Labeling Manager should un-assign the current task and assign the High Priority task to the labeler(s) using the Assignments tab in the Labeling Manager Console.
  + The Labelers X Tasks grid in the Console (shown below) allows the Labeling Manager to see all the labeler assignments to various tasks and to fine-tune which labeler works on which task in order to support the task priorities.
* **Setting Validation Filters** -



* **Rejected documents** - Rejected state of the document along with the Reason code [e.g. different doc-type, forged, glare, blurry, edge cut off, etc] is captured in the Document JSON files ([TextChange](https://cloud.google.com/document-ai/docs/reference/rest/v1beta3/Document#textchange) construct for all changes, [HumanReview](https://cloud.google.com/document-ai/docs/reference/rest/v1beta3/Document#humanreview) for reject reason), output in the configured GCS bucket. It is recommended to use this metadata to separate the rejected documents.
* **Handling large document ingest volume** - Current quotas (<https://cloud.google.com/document-ai/quotas> T) limit 600 online document requests/minute and upto 5 concurrent offline (batch) requests, each handling upto 50 documents. This is expected to handle 36000 online requests/hour and upto 15000 offline (batch) document requests/hour [since each batch handles upto 50 documents]. If higher volumes are required, please contact Support.
* **Managing human validation costs and time** - Human validation can get expensive, whether you use your own labelers or Google labelers. The time spent per page scales with the # of fields reviewed. Google Workforce (in Private Preview) is billed per label reviewed. You can control your review effort and cost by limiting HITL to the important fields that need to be verified and corrected. To do so, use the **Label-level Validation filters** in the Human Review configuration screen.
* **Single Sign-on access with Enterprise credentials** - the Labeler Workbench and Labeling Manager Console support Google Workforce or Gmail credentials. If single sign-on with enterprise credentials is required, please create Basic Google Workforce accounts for your labeler workforce. Google Workforce supports SAML and you can configure your SSO provider to sign onto the Labeler Workbench or Labeling Manager Console with enterprise credentials.

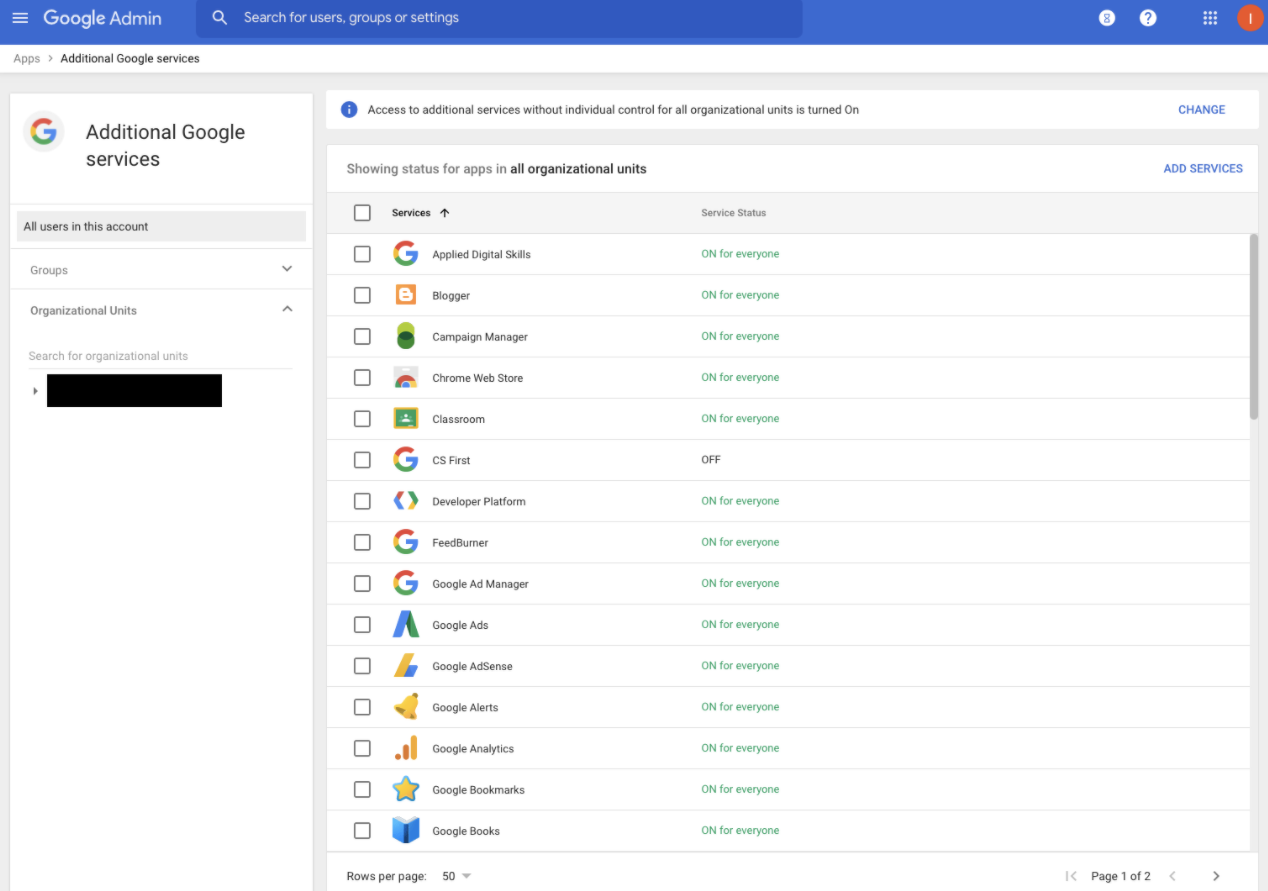
## Known Limitations and Workarounds

Following are some known limitations and possible work-arouunds. We are working on addressing these limitations in the near-term.

* **VPC-SC** - Customers with VPC-SC activated in their HITL project may not be able to use Google labelers for HITL Review.
* **Post-review Latency** - After a document is reviewed, it may take up to 30 minutes for it to be deposited in the pre-configured GCS folder.
* **10 page limit** - HITL review is limited to 10 pages for Invoices. Invoices larger than 10 pages are not sent for HITL review.
* **Single-labeler workflow** - The current workflow tasks are limited to a single labeler review. If the document needs to be reviewed by multiple reviewers (for QA, fraud prevention etc), please upload the reviewed documents to a second processor in the mode "Skip processor, send all docs for HITL Review", so they are reviewed.
* **Edit Labeling Managers of Labeling pool** - In BYOL tasks, labeling managers cannot be added or removed to a labeling pool once it’s created. Workaround: When creating a new labeling pool,
  + Please assign more than 1 labeling managers, so that if one leaves before the task completes, the other(s) can manage the pool and task assignments.
  + Individuals that need to view the labeler or task analytics dashboards in the Labeling Manager UI should be added as Pool Managers when creating the pool.
* **Canceling a Task** - Once started, a task cannot be cancelled. The work-around is for the Labeling Manager to pause the task or unassign labelers from the task in the Labeling Manager Console.
* **Assigning a Task to an entire Labeler Pool in a single click** - isn’t currently supported but will be supported shortly. The work-around is to select all the labelers and assigning the task to them.



* BYOL Labelers with G-Suite accounts - If BYOL labeling managers and labelers have G-Suite accounts, the GSuite Admin may need to enable “Other Services” at the bottom of the Google Services page (screenshots below), in order to enable access to the Labeling Manager Console and Labeler Workbench.



## 

# Quickstart: Configuring a HITL Document Review Task

[LINK - [https://cloud.google.com/document-ai/docs/hitl/quickstart]](https://cloud.google.com/document-ai/docs/hitl)

## Customer Steps

### Step 0: Enable the API Platform (Unified) API

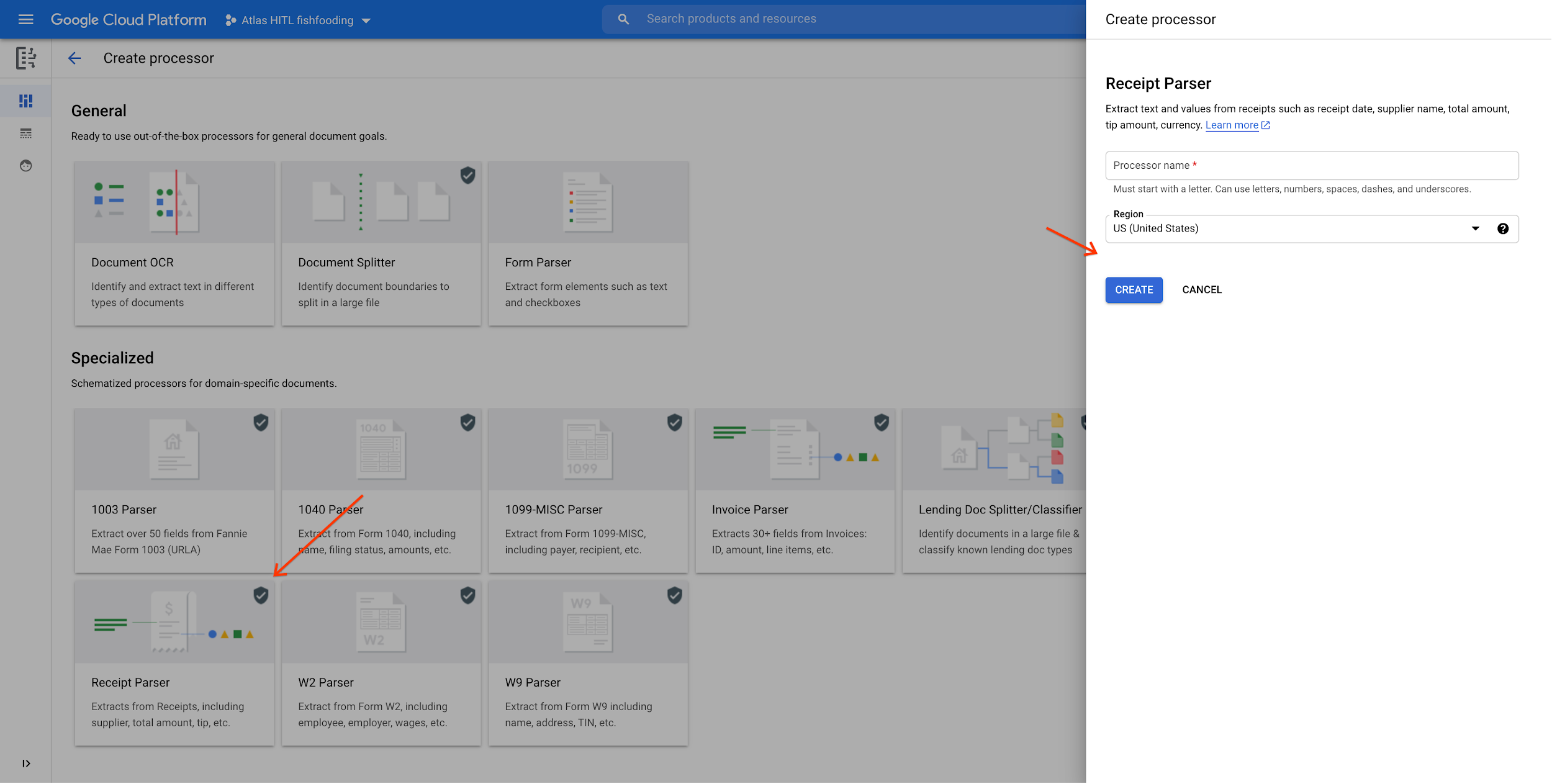
This step should already have been completed in order to set up any DocAI Processor. See Document AI documentation.

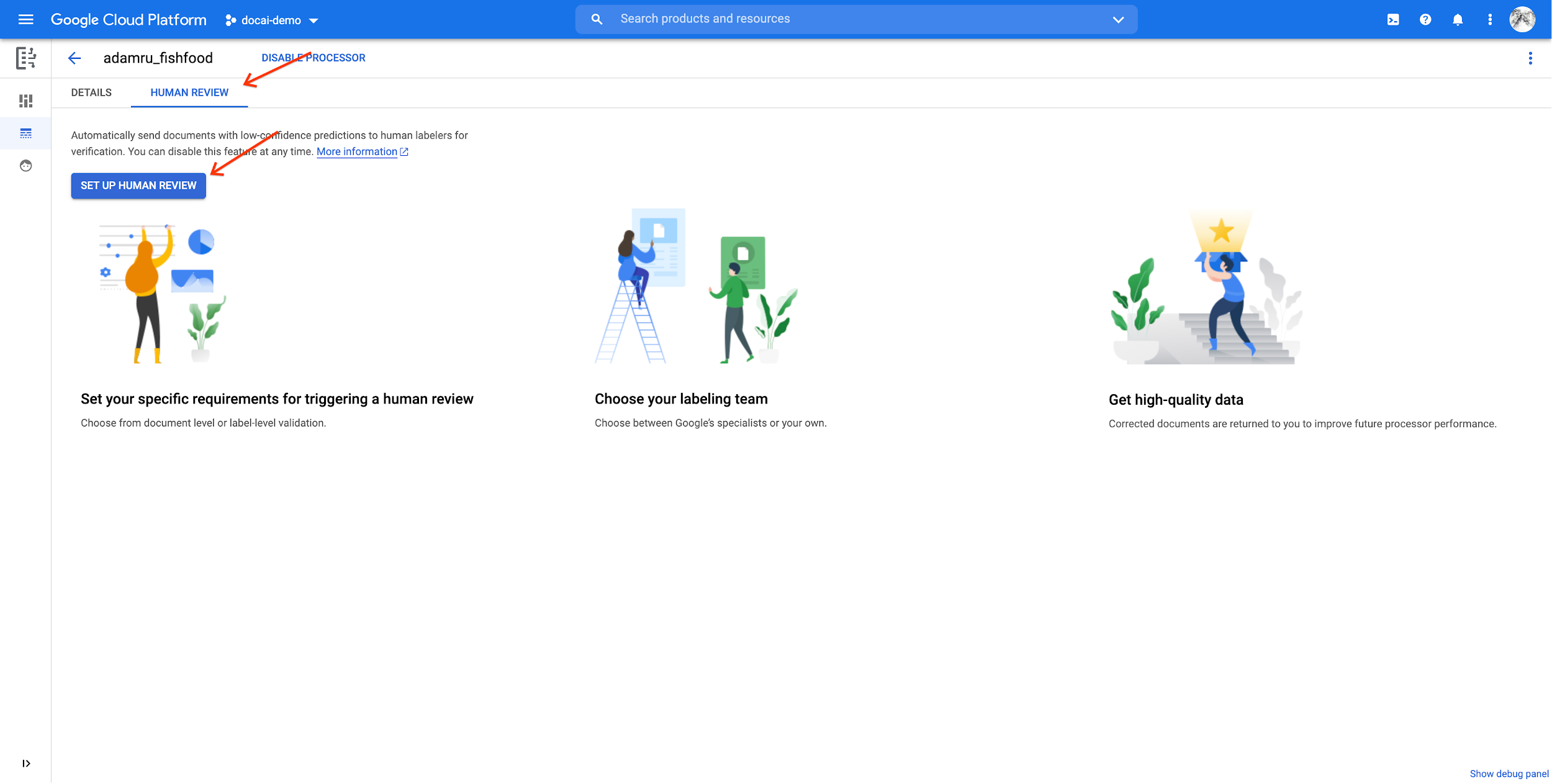
### Step 1: Create a processor and set up human review

NOTE: Currently, the HITL Review workflow is supported for Invoices and Receipts, and support for HITL on additional processors is planned.

Each processor creates a “HITL task” that is assigned to a pool of human labelers (called “Labeler Pool”) that review the documents processed by the processor. Once processed, these documents are queued up in the task for HITL review by the assigned Labeler Pool.

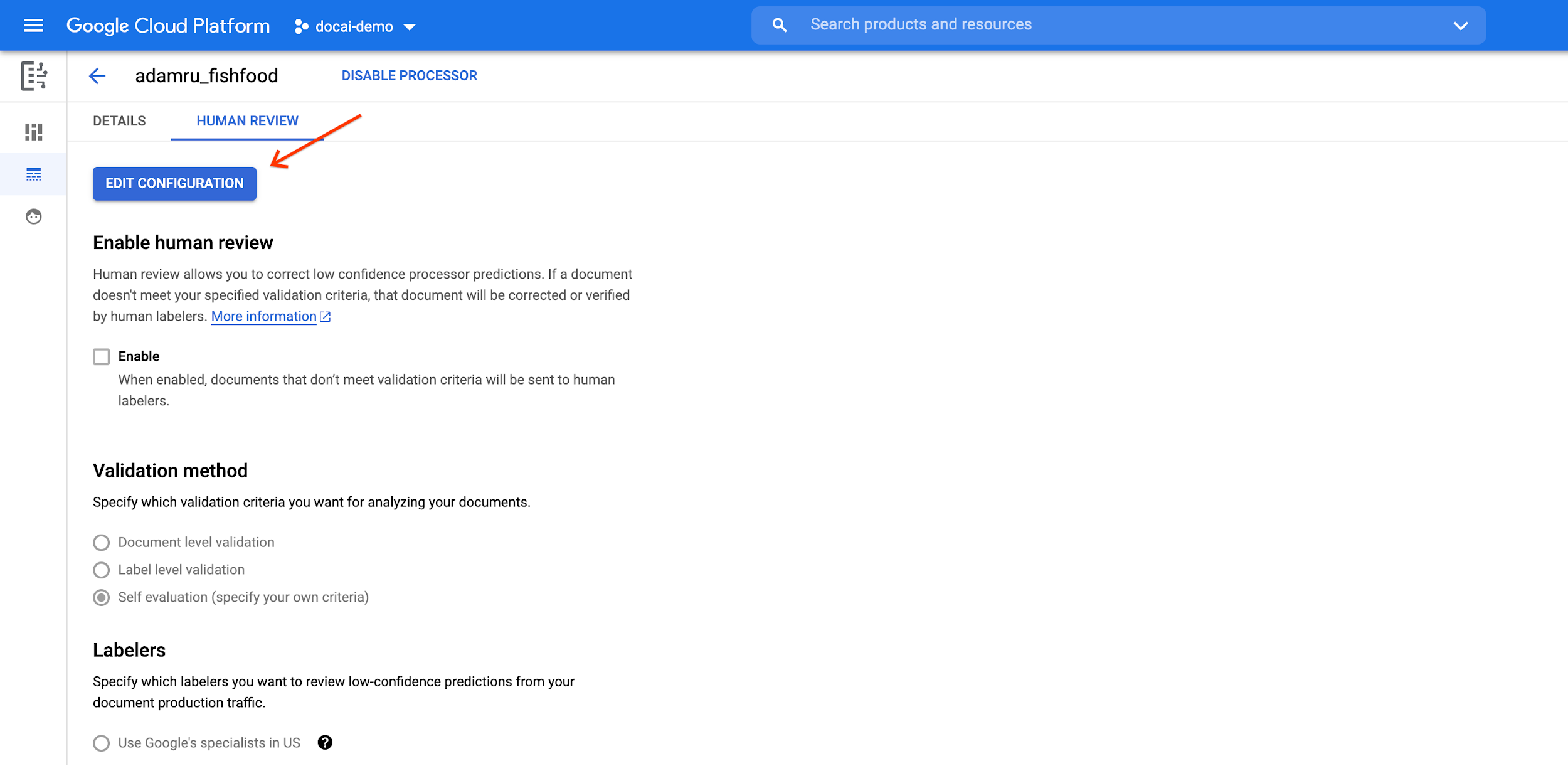
* 1. Go to the Document AI [“Overview”](https://screenshot.googleplex.com/6ngazha2h65Vcrt.png) page and create a processor.

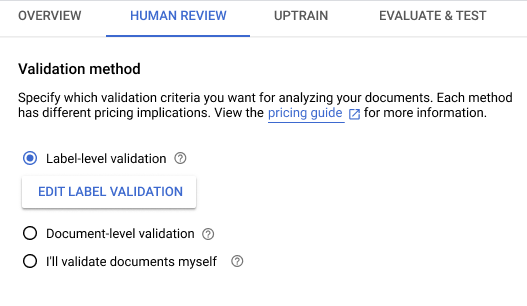
For example, a Receipt Parser (note: Invoice Parser is illustrated in following screenshots)  


* 1. After creation go to the “human review” tab and click the setup button.  
     
  2. You will see a “Initializing human review” page:

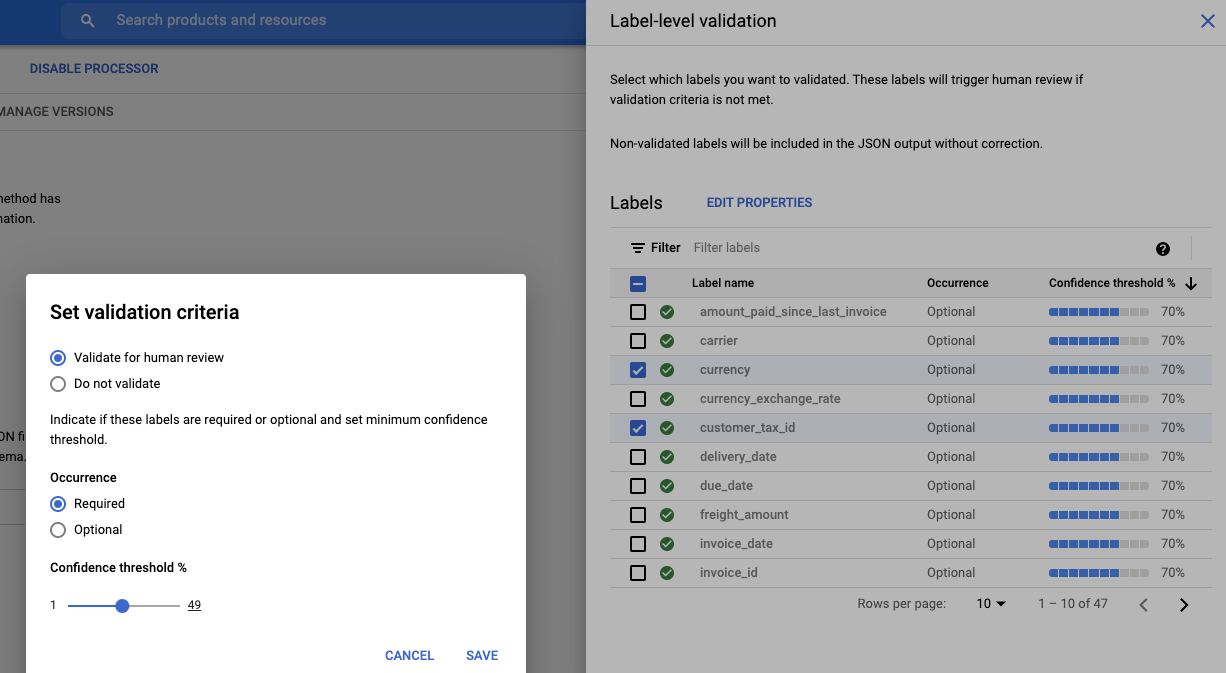
  
After a few minutes (for the first time, while resources are being set up), it will auto refresh to [edit page](https://screenshot.googleplex.com/9ugwbhTUwjf8nwu.png) after initialization finished.

### Step 2. Configure Human Review

* 1. Click the Edit Configuration button:  
     
  2. Click the **Enable** check box. It must be checked (can be done later) to route documents to Human Review.
  3. Scroll down to [Validation Method section](https://screenshot.googleplex.com/7a4FuYNKnPf8aSt.png),

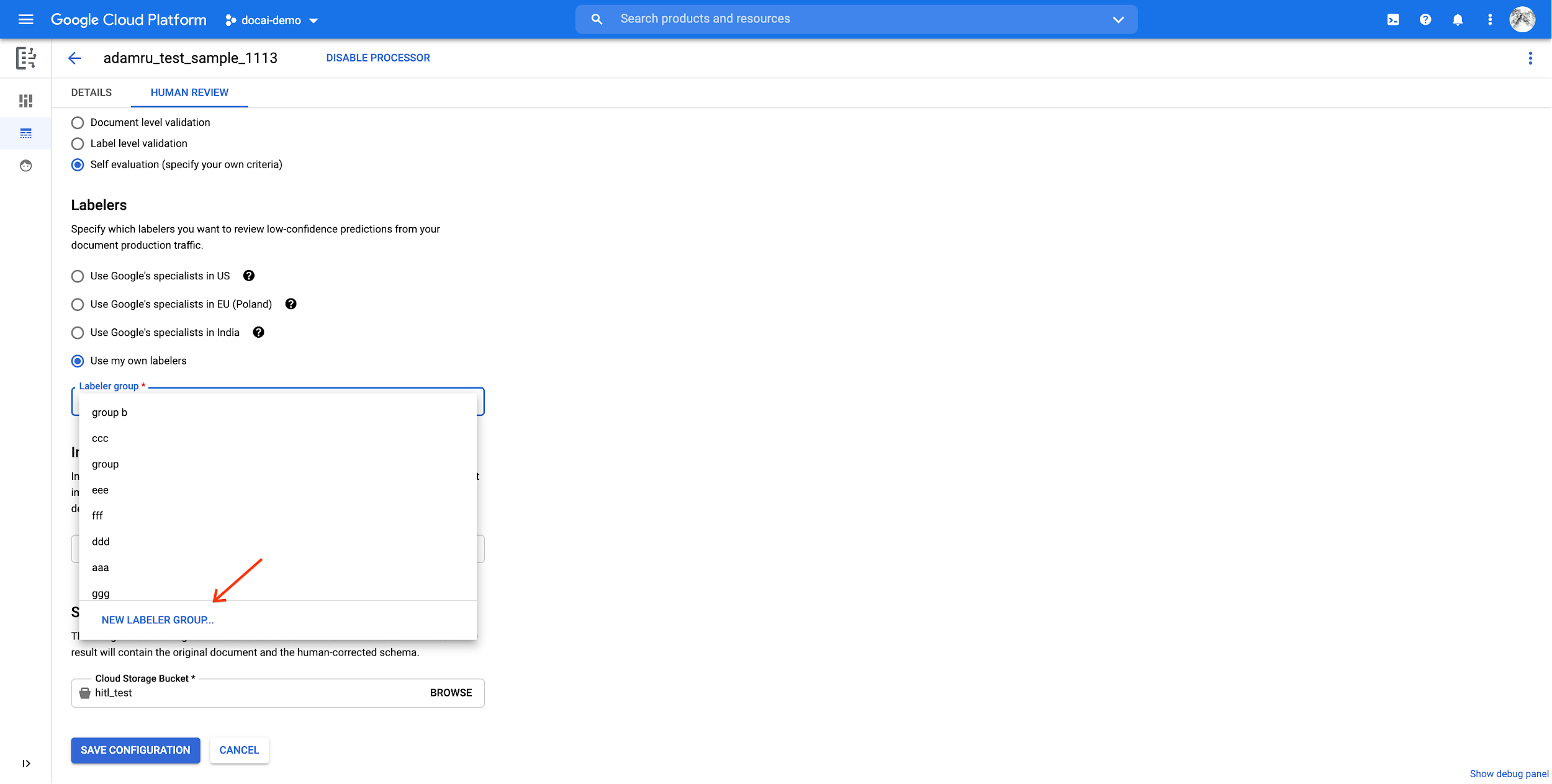
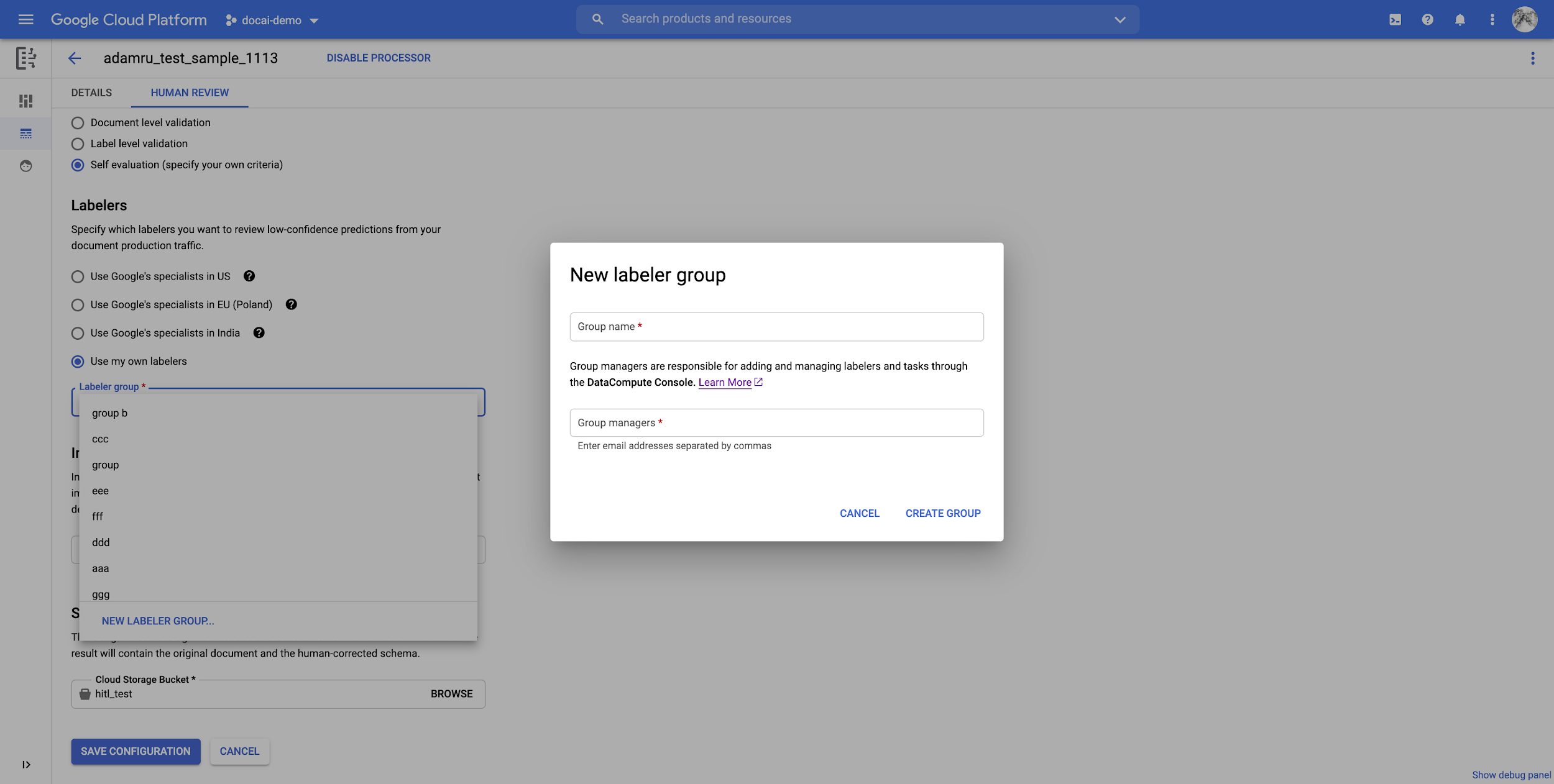
  
  
Select “**Label-level validation**”, and Edit Label Validation.

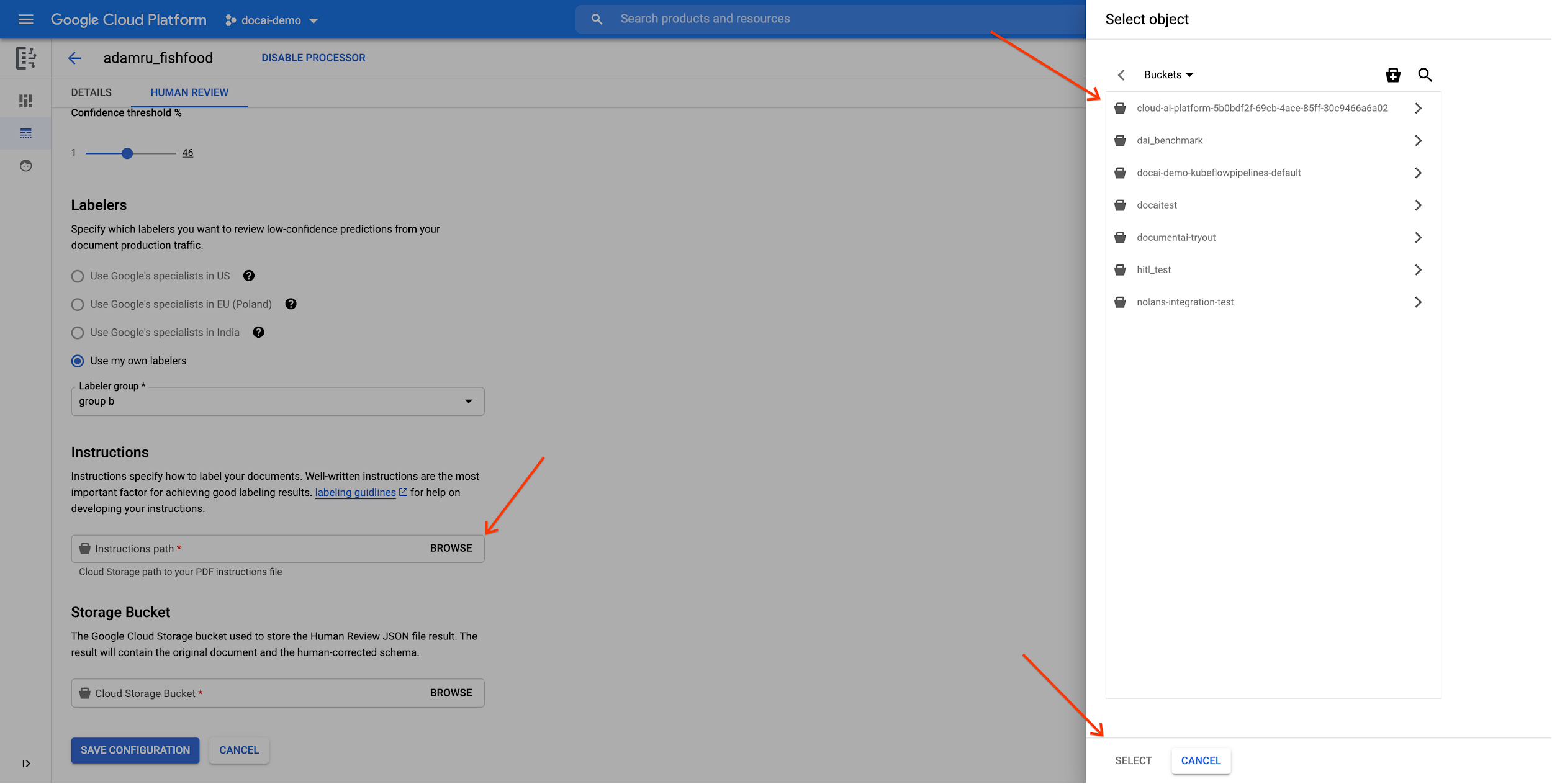
* + 1. Select the fields required for review -> Edit Properties
    2. Select Validate for Human Review
    3. If any fields are Required in the extracted results, please toggle Occurrence to Required
    4. and to change the threshold with the slider, if desired.
    5. Adjust Confidence thresholds as required

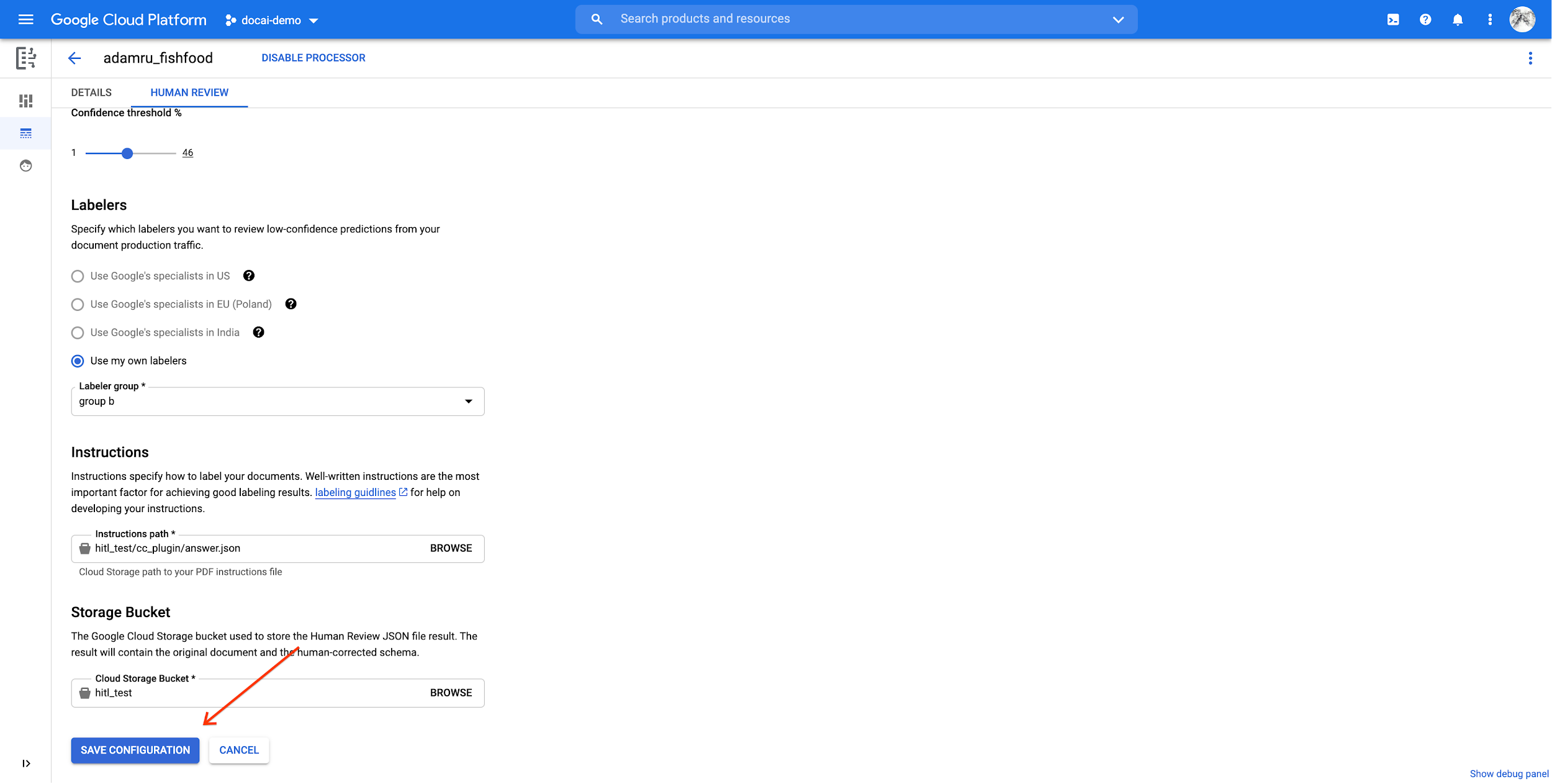


* 1. Select **Bring-your-own-labeler: Use my own labelers** (default).

NOTE: The [**Google HITL workforce**](https://cloud.google.com/document-ai/docs/hitl/google-workforce) **option** is in Private Preview and is limited to processing only non-PII data. If you’re interested, please follow instructions [here](https://cloud.google.com/document-ai/docs/hitl/google-workforce).

* 1. Scroll down to the **Labelers** section, click the dropdown arrow next to the Labeler Group field, then select click “NEW LABELER GROUP...”   
     
  2. Create a labeler group with the dialog, supplying a group name and an email address(es) for manager(s):  
     Click CREATE GROUP. It will take several seconds.
  3. Select it as your **Labeler group** (it is by default).
  4. Prepare a PDF file as your labelling instructions. Refer [here](https://cloud.google.com/ai-platform/data-labeling/docs/instructions) for how to design good instructions. Upload it to a bucket in your project.
  5. Scroll down to the **Instructions** section. In the **Instructions location** field, click BROWSE and select an pdf instruction. Click SELECT



* 1. Scroll down to the **Results** section. In the **Results location** field select a bucket for the Human Review output.
  2. Click the SAVE CONFIGURATION button.  
     The page will refresh with changes

## Labeling Manager Steps

### Step 3: Add Labelers

* 1. You should receive an email, like:

From: AI Platform <noreply-aiplatform@google.com>

Subject: Added to AI Platform SpecialistPool Test Labeler Group

Inbox

Hello AI Platform Customer,

You are added as a manager to specialist pool

cloudml\_data\_specialists\_us\_central1\_785484070008756xxxx !

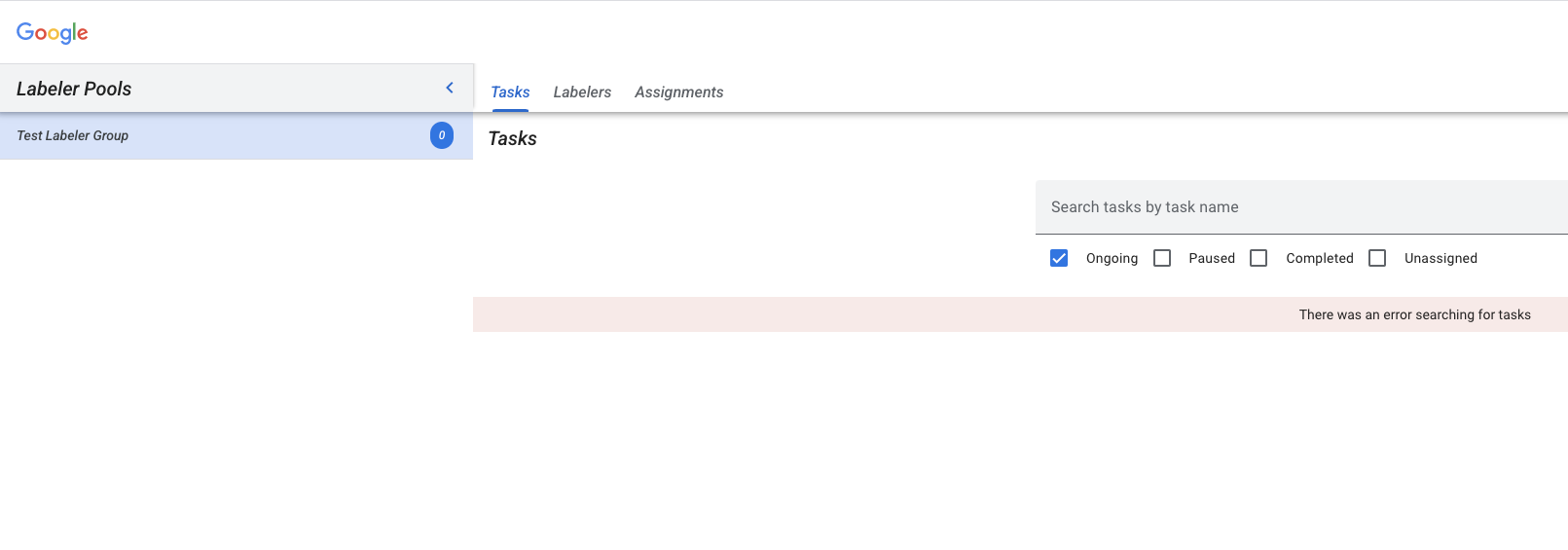
To view details, visit Manager console

[https://datacompute.google.com/cm/](https://datacompute.google.com/cm)

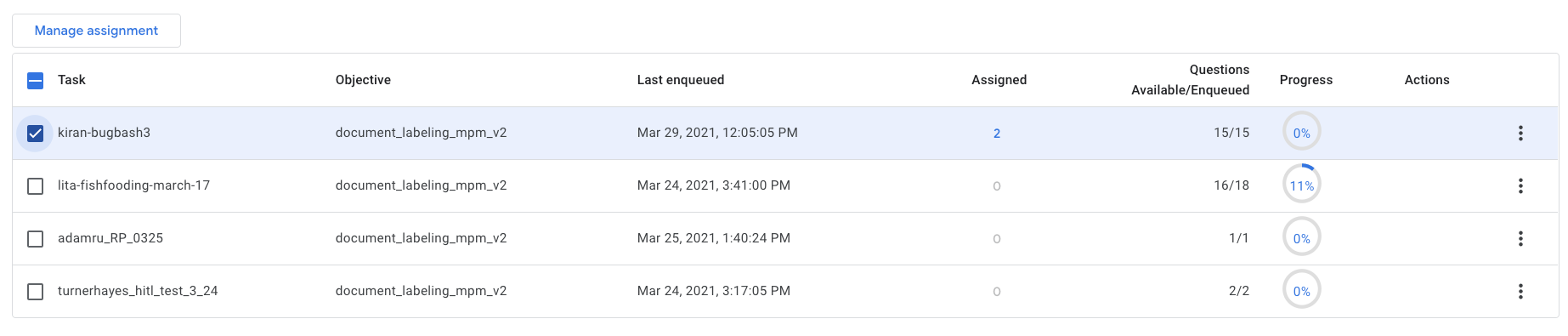
Sincerely,

The Google Cloud AI Team

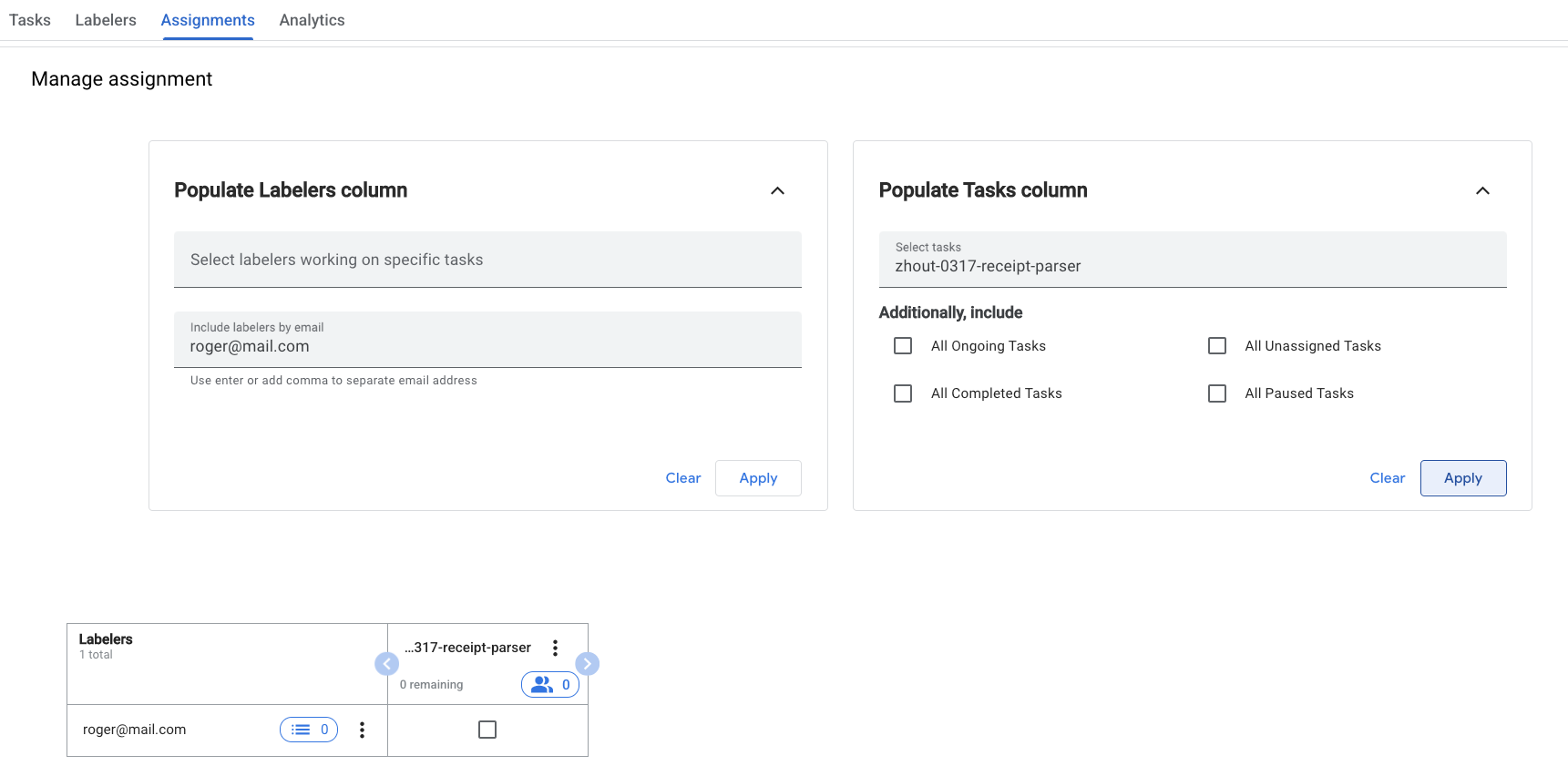
* 1. Click on the link in the email (or cut and paste into your browser) to navigate to the Manager page:



* 1. The task created by the newly created Processor in Step 1 has the same name as the Processor (e.g. “Department A Invoices”) and is listed in the Tasks tab.



* 1. Click on Labelers tab
  2. Click on Add Labeler, and add at least one labeler (Email). A manager can be a labeler, so you can add yourself.
  3. For a newly created Labeler Group, the task needs to be assigned to them explicitly in the Assignments tab. That is, adding the Labeler Group in Step 3e doesn’t automatically assign the task to these new labelers. If the Labeler Group in Step 3e has already provisioned labelers, the pool is automatically assigned to the task.



### Step 4: Post an Initial Document to Review

1. Send a document for extraction thru the created Processor. Refer to [How-To Guides](https://cloud.google.com/document-ai/docs/how-to) as needed. **Both sync (process method) and batch (processBatch) calls currently support Human Review routing for Invoice and Receipt Parsers.**

For test purposes, using a document with known poor extraction would be best to trigger the Human Review by the confidence score being below the threshold- or set the threshold to 100%. Document limits are generally 5 pages, 20MB max, but check the specific Processor limits.

NOTE: Uploading a document [via the GCP console UI](https://cloud.google.com/document-ai/docs/quickstart-console) to test the processor will **not trigger** a Review process.

**DEVELOPER NOTES:**

There is API Access to track each document routed to Human Review.

As part of the [response from Processors](https://cloud.google.com/document-ai/docs/reference/rest/v1beta3/projects.locations.processors/process), there is a humanReviewOperation string that is the Operation ID (job name) of the document within Human Review. This (Long Running) Operation (also referred to as a LRO) can be queried for status. **There is no method to see all such LRO’s, so retaining the Operation ID is essential.**

You can query the status of a document routed to Human Review with the [projects.locations.operations.get](https://cloud.google.com/document-ai/docs/reference/rest/v1beta3/projects.locations.operations/get) method

NOTE: The [Sync process method](https://cloud.google.com/document-ai/docs/reference/rest/v1beta3/projects.locations.processors/batchProcess#BatchInputConfig) includes a skipHumanReview

boolean: Whether the Human Review feature should be skipped for this request. Default to false.

1. You can force a document to be Human Reviewed with this method (note that it is Processor-specific):

[projects.locations.processors.humanReviewConfig.reviewDocument](https://cloud.google.com/document-ai/docs/reference/rest/v1beta3/projects.locations.processors.humanReviewConfig/reviewDocument)

Using the reviewDocument method requires the Human Review ENABLED checkbox to be checked. Using this API will not override the config settings.

### Step 5: Review the document

Manager Email:

If the document triggers the human review process, but there are no Labelers in the pool, the manager will get an [email](https://screenshot.googleplex.com/7MsSM7nNe4fq496.png),

Hello AI Platform Customer,

There is a running task for Project with Job ID 404298342016955xxxx,

Labeling Task ID cloudml\_data\_us\_central1\_prod\_11b273b4dabdxxxx,

Display name

datalabeling-job-8005676103232389120-for-hitl-340697311810578xxxx,

replication count 1

Please see attachment for instruction.

Please add labelers and have them finish the task at Labeler console

<https://datacompute.google.com/w/cloudml_data_specialists_us_central1_7854840700087566336>.

To view details, visit Manager console

<https://datacompute.google.com/cm/cloudml_data_specialists_us_central1_7854840700087566336/tasks>.

Sincerely,

The Google Cloud AI Team

These links are for the Labeling Manager to allocate labeling work.

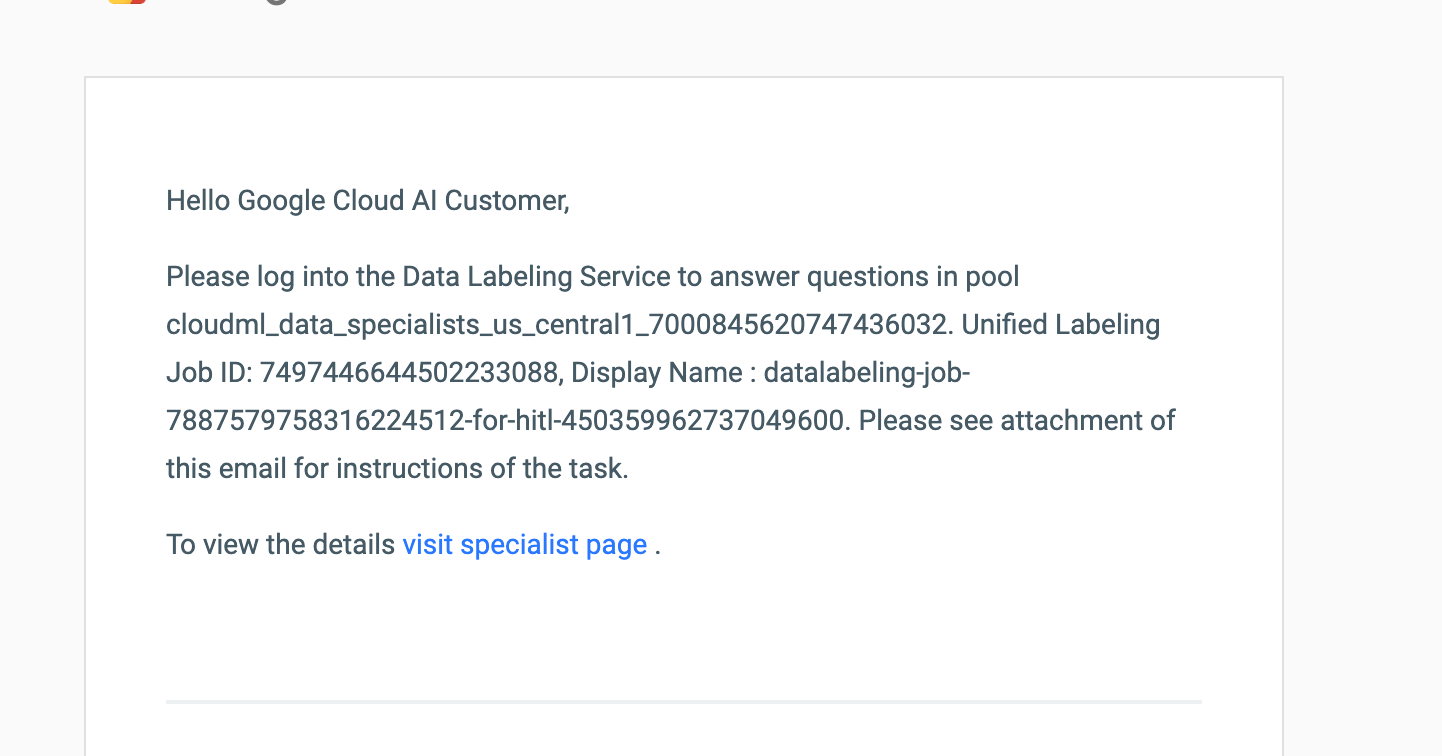
1. Labeller Email  
   If the labelers aren’t enrolled in the Labeling Group and are newly added in Step 3d, the Labeling Manager needs to send a similar email with instructions to the Labelers added.

NOTE: The Manager may manually re-allocate routing of future new documents to labelling pools (tasks). See “Manage Human Review Queues”.

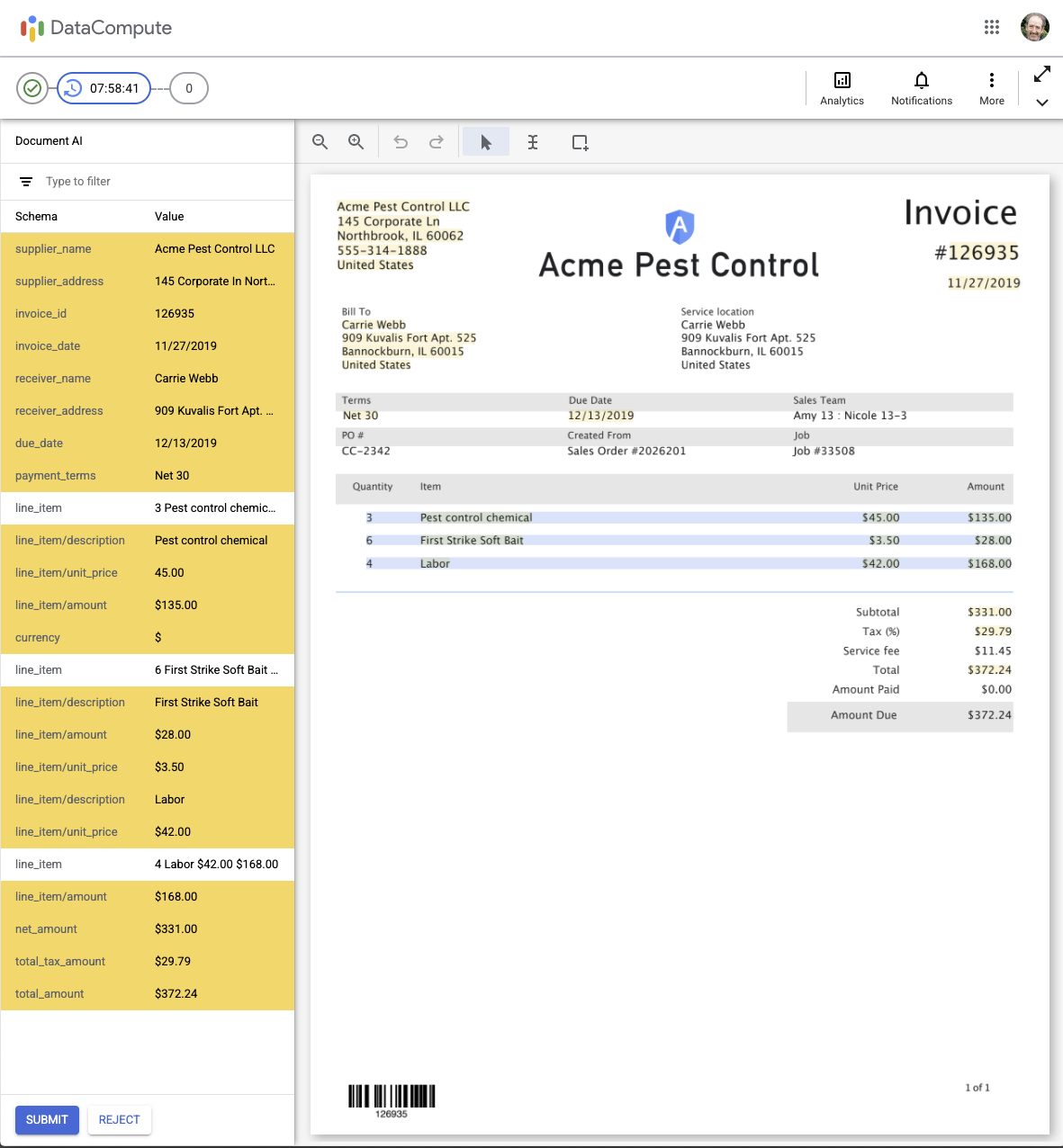
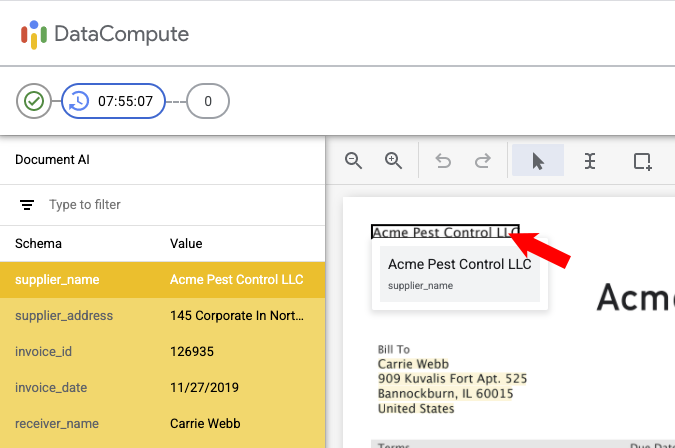
# Labeller Steps

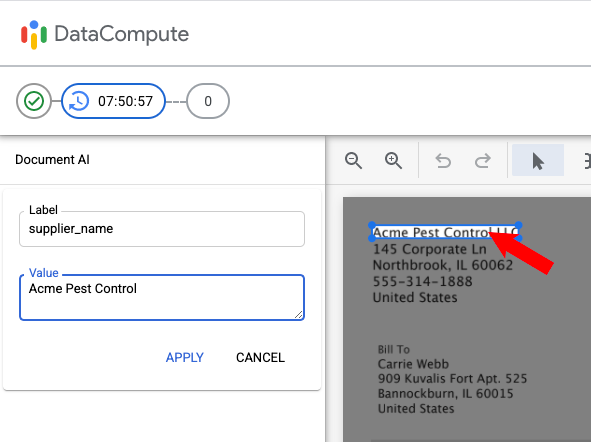
### Step 6: Labeller Procedure

1. Labeller Email

If a labeler is already enrolled to a Labeling Group in Step 2e, the system will auto-allocate, and they will get this sort of email:  


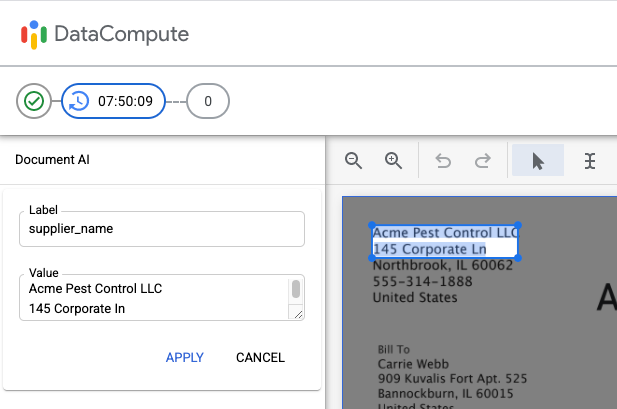
The link in this email is the mechanism for the selected Labeler to do labeling work on the document.

1. Labelling Actions:  
   When a Labeller clicks the link to do the labeling, they will see the Interactive AI Human Review User Interface shown below. The highlighted (in yellow) rows indicate the rows that are failing the validation (that is, they are below the configured confidence score threshold) and need review.  
   1. Review  
      By hovering the cursor over a recognized text field, the recognized text and (In smaller font below it) the name of the field (schema) is shown. The row is also highlighted in the left panel.  
        
      The user may zoom in/out with the magnifier icons at the top left of the toolbar.   
        
      PLEASE NOTE TO USERS: **To scroll horizontally, hold down the shift key and use the scroll wheel on the mouse.** This is not documented elsewhere.  
      
   2. Change Field:

By (left) clicking on the field, you can edit it:

You may change the value or label in the left panel. Then click **Apply**.

OR, by clicking on one of the resize corner icons on the document image, you may resize the bounding box around the document text:

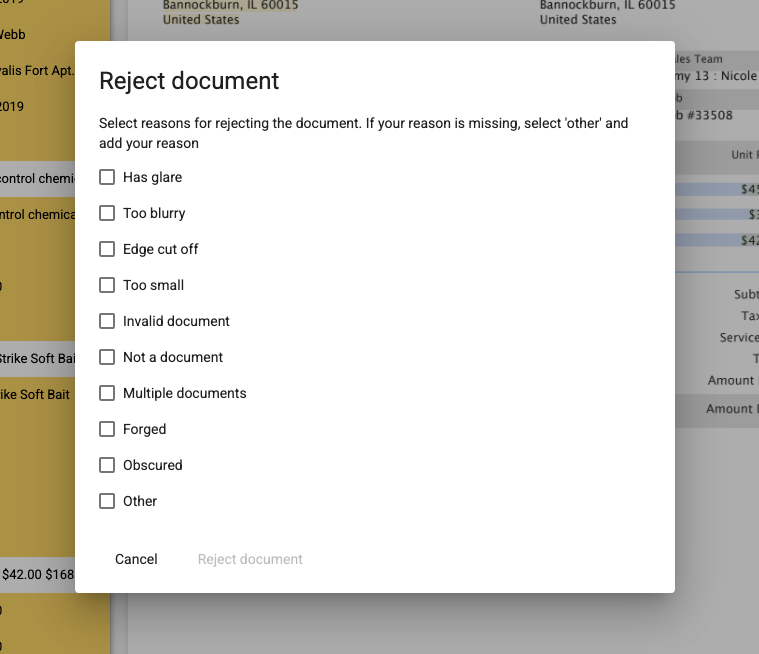


As shown, if different text is selected, it will change the text in the field’s Value in the left panel. You can then further edit this text if needed. Then click **Apply**.

Be sure to review all pages, using the page control in the upper right corner.

After making all needed label changes, click **SUBMIT** (at bottom left) result json will be saved to the “Results location” folder.

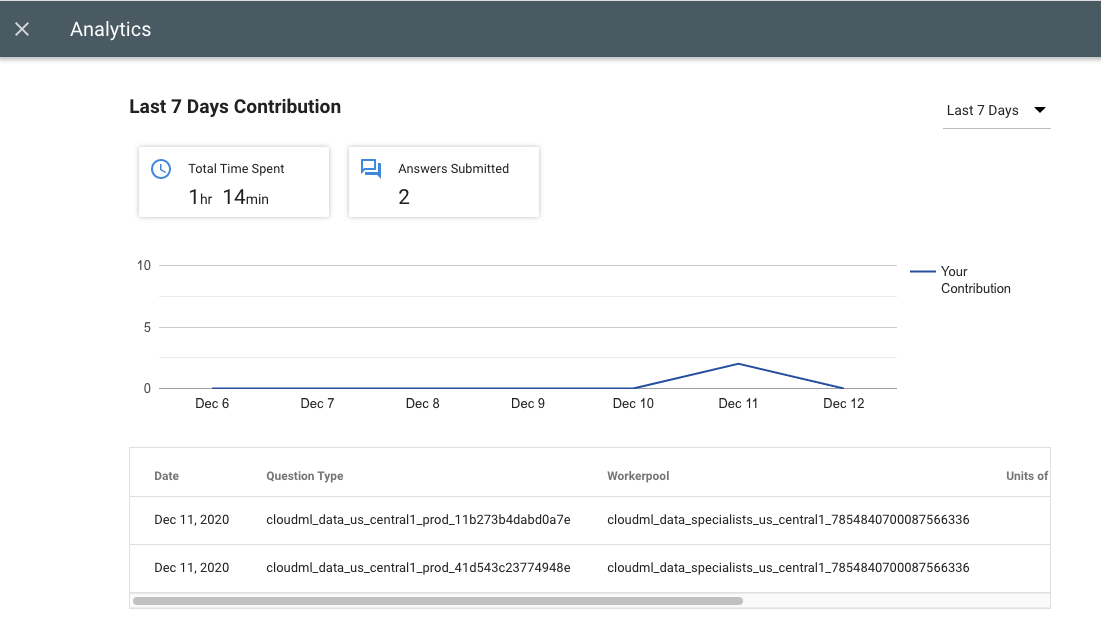
If the document cannot be satisfactorily labeled, you may click **REJECT** (at bottom left), and then select a reason for rejecting the document:

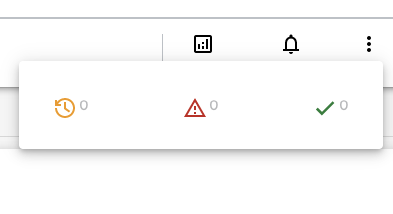


Select a reason, then click **Reject document**.

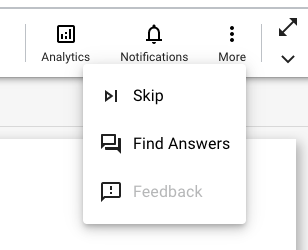
Labellers may also click on the following icons in the upper right corner:

**Analytics** - to show how many documents they have reviewed (Answers) and their total time:



**Notifications** - To show any Notifications they have gotten:  
 

**More (3 Dots)** - To **Skip** the document, **Find Answers**, or provide **Feedback** (if enabled):



**Toggle Title bar (Diagonal Arrows)** - This will hide (or show) the DataCompute title bar for more room.

### Step 6: [Manager] Review Document from Cloud Storage

If you want to check the document, there are two options:

Option 1:

1. Locate the Cloud Storage from previous configured location
2. Retrieve and download the documents

Option 2:

1. Remember the URI for retrieve the document from Prediction response
2. Call URI to retrieve the reviewed document

Developer’s Note: Human Corrections will be recorded in the [TextChanges](https://cloud.google.com/document-ai/docs/reference/rest/v1beta3/Document#textchange) construct of the Document JSON.

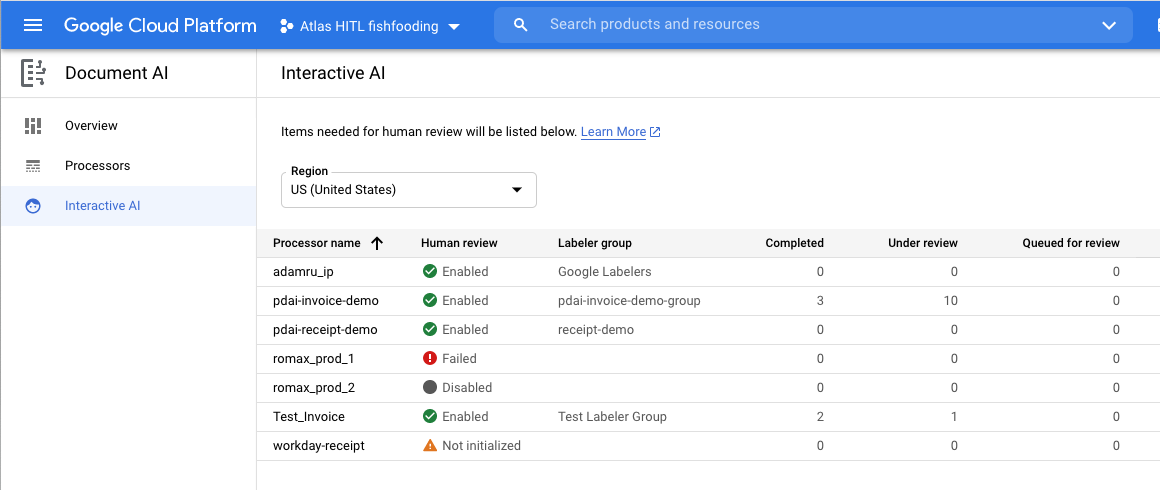
### Step 7: [Manager] View HITL Review Analytics by Task and by Labeler

<Add Analytics screenshots>

# **Labeling Manager:** Manage your labeling tasks in DataCompute

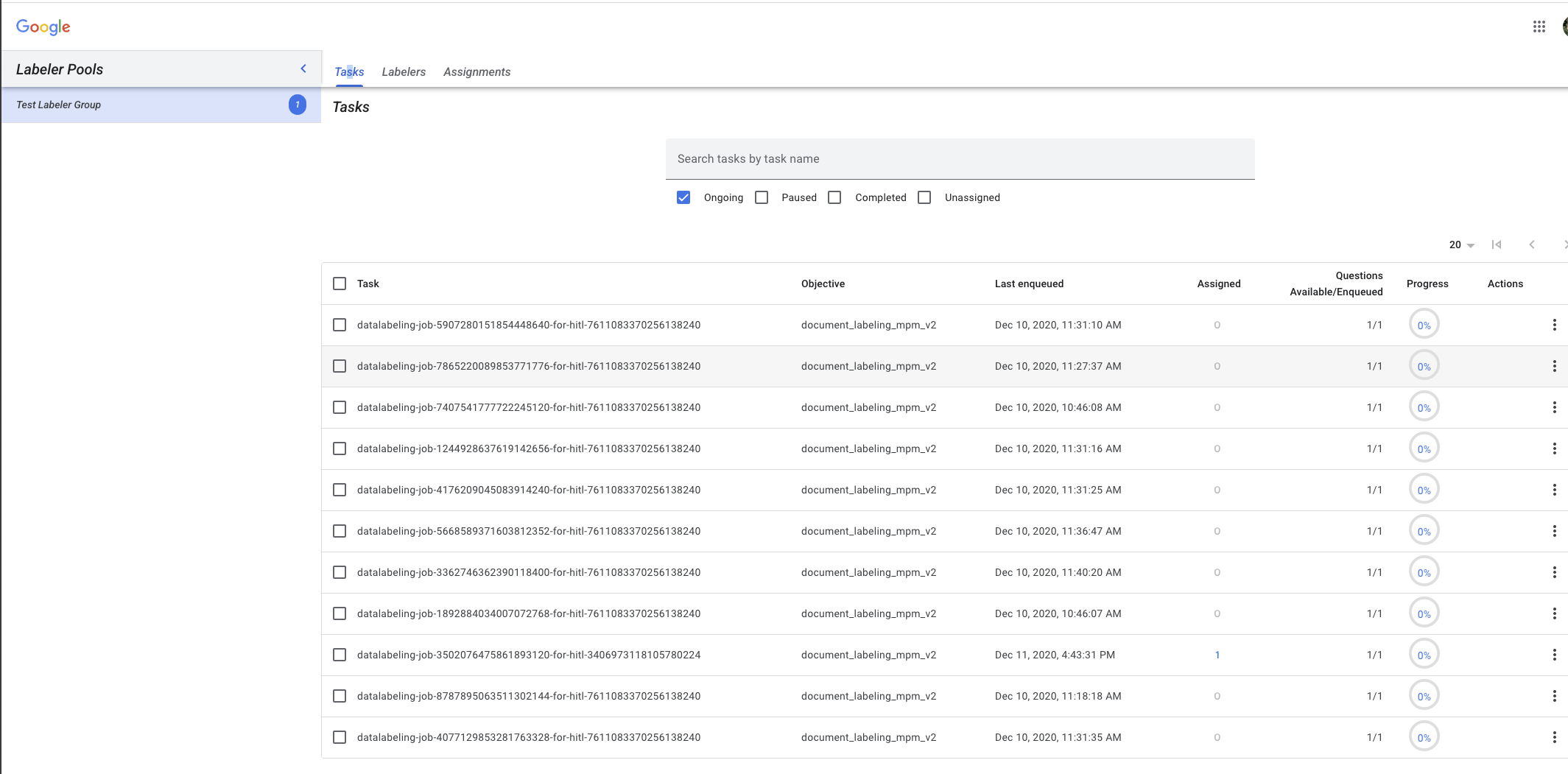
## Manage Human Review Queues

[Manager]: In the GCP console under Document AI, Interactive AI, you can see the counts by project for the status of documents:



From the Manager console (accessible from the link sent in email) the Manager can manage Tasks, Labelers and Assignments:

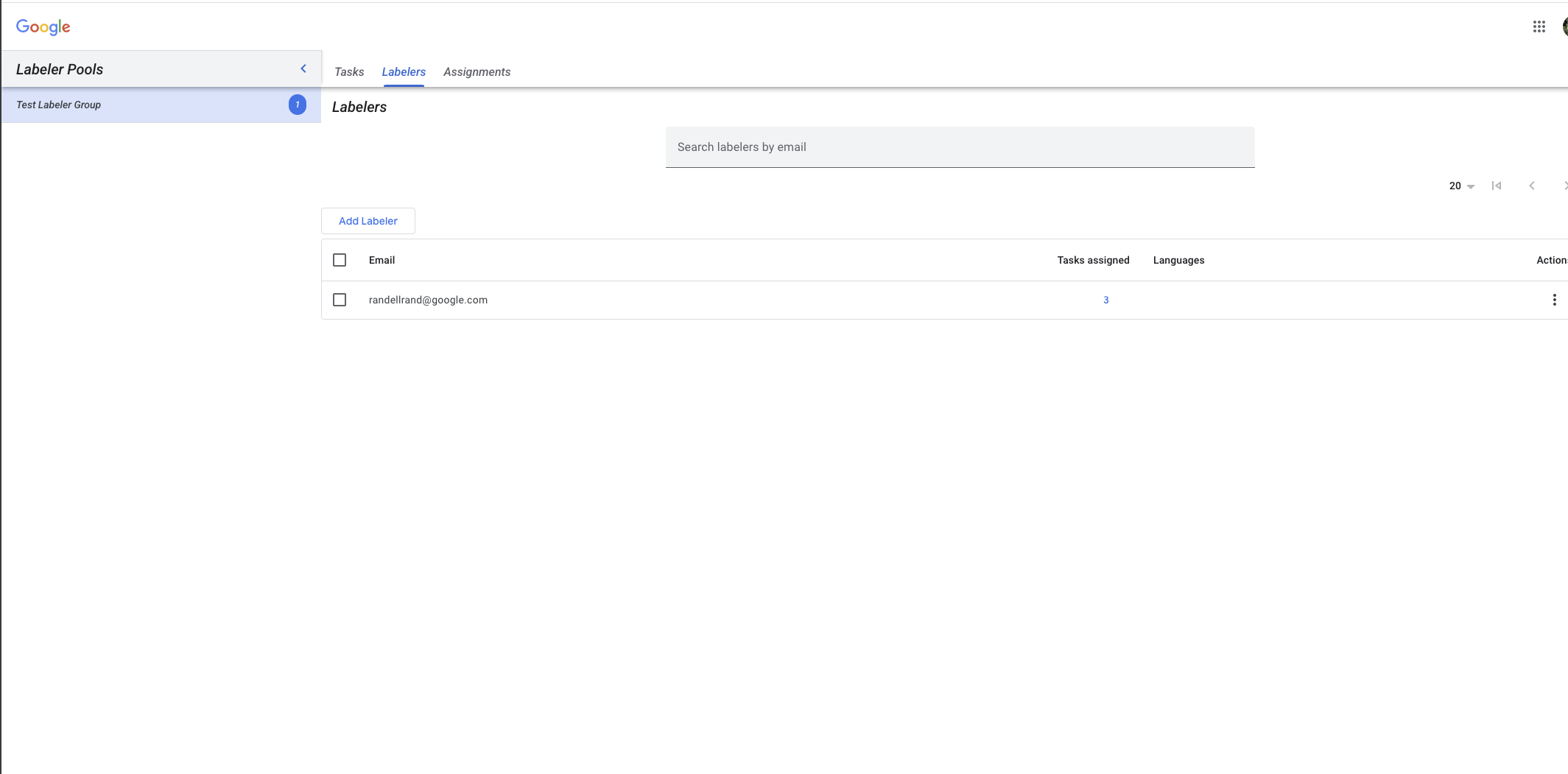
The **Tasks** menu item lists tasks:



By clicking on the 3 dots to the right of entry, you can:

* Pause
* Perform Assignment

The **Labelers** menu item lists Labelers:

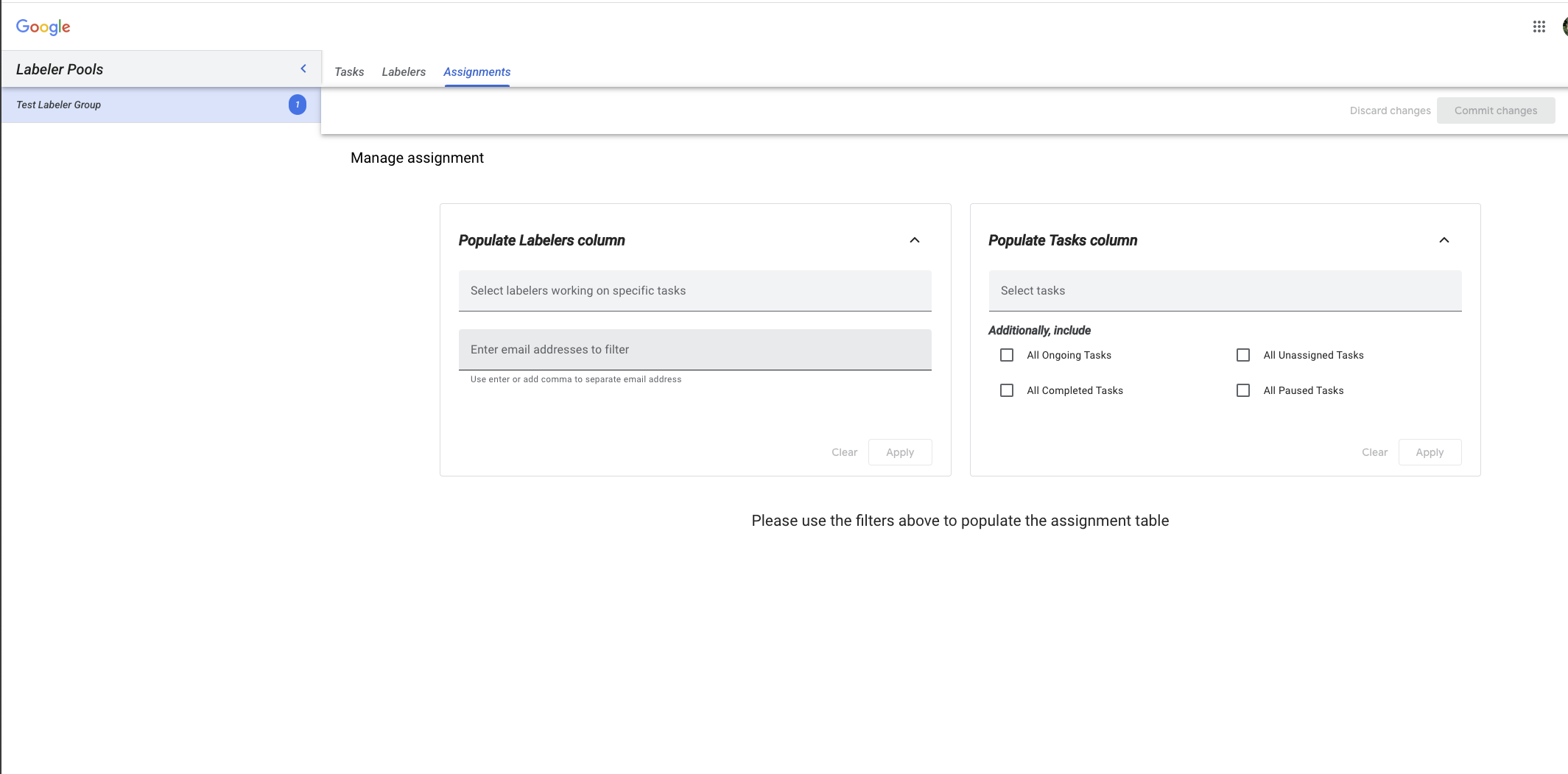


With the **Add Labeler** button, you can add Labelers.

By clicking on the 3 dots to the right of entry, you can:

* Delete
* Perform Assignment

The **Assignments** menu item manages Assignments:



## 

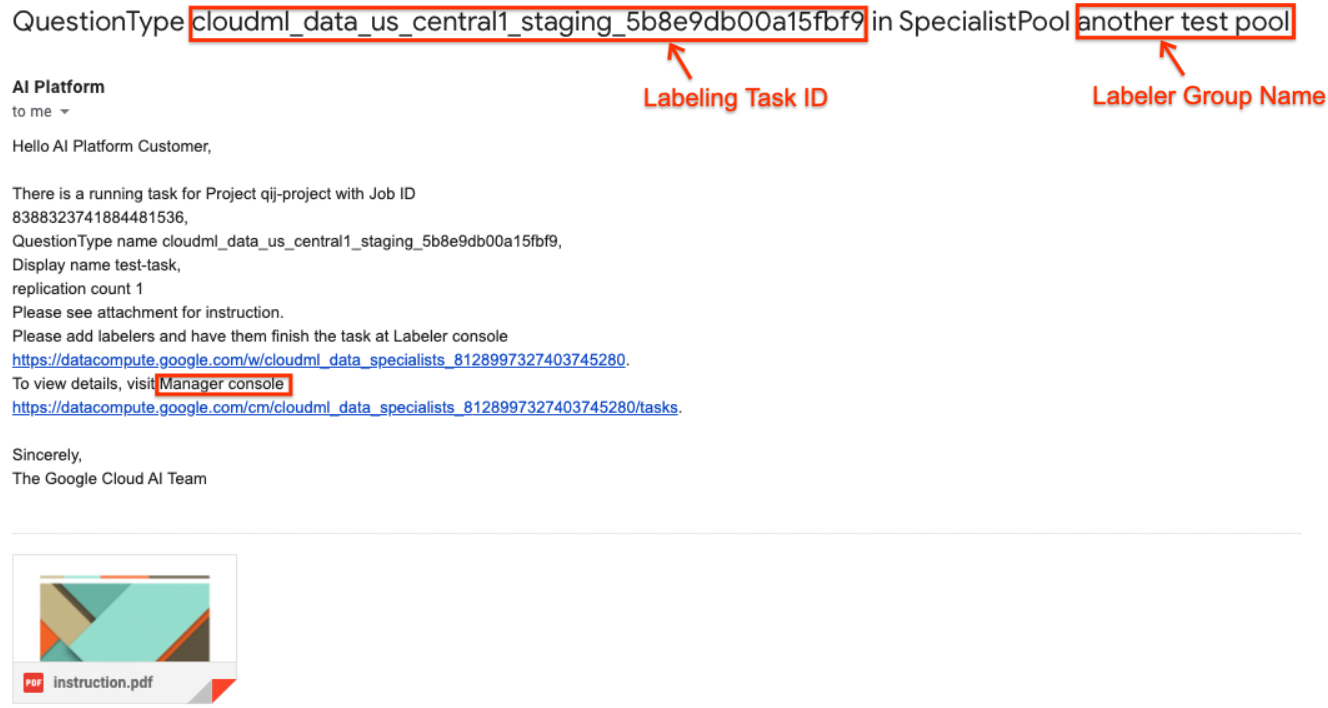
There are two kinds of roles in a labeler group: labelers and labeler managers.

* **Labelers** are the actual human reviewers for your task.
* **Labeler managers** manage the labelers. They can track the task progress, as well as add/remove labelers to/from a task. A typical labeler group usually has multiple labelers and at least one manager.

Both labelers and managers use the DataCompute console to handle the labeling tasks.

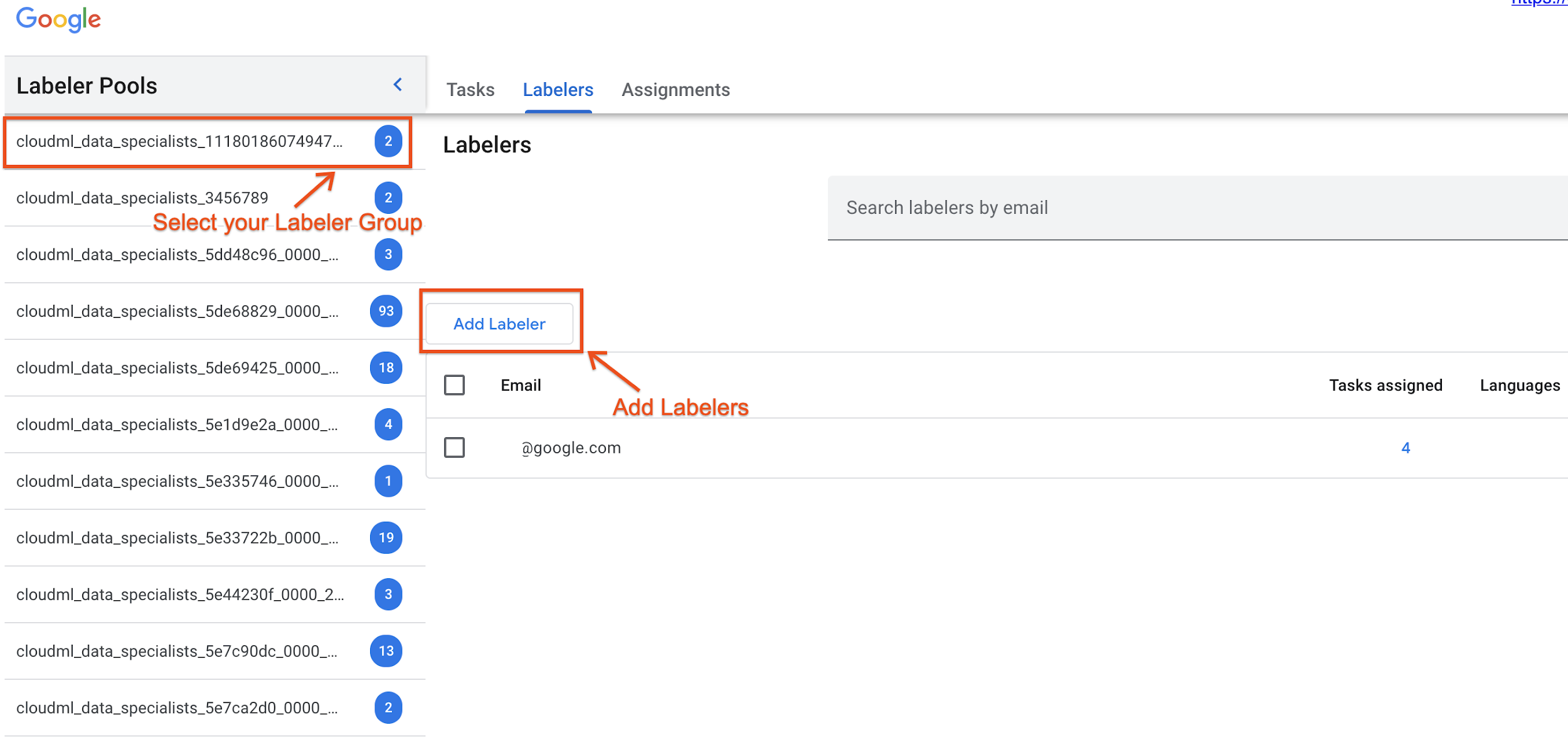
## Receive Email Notifications

Once a labeling group is created by the customer, the labeler managers will receive an email similar to the one listed below. They can login to the “Manager console” to manage labelers for the task.



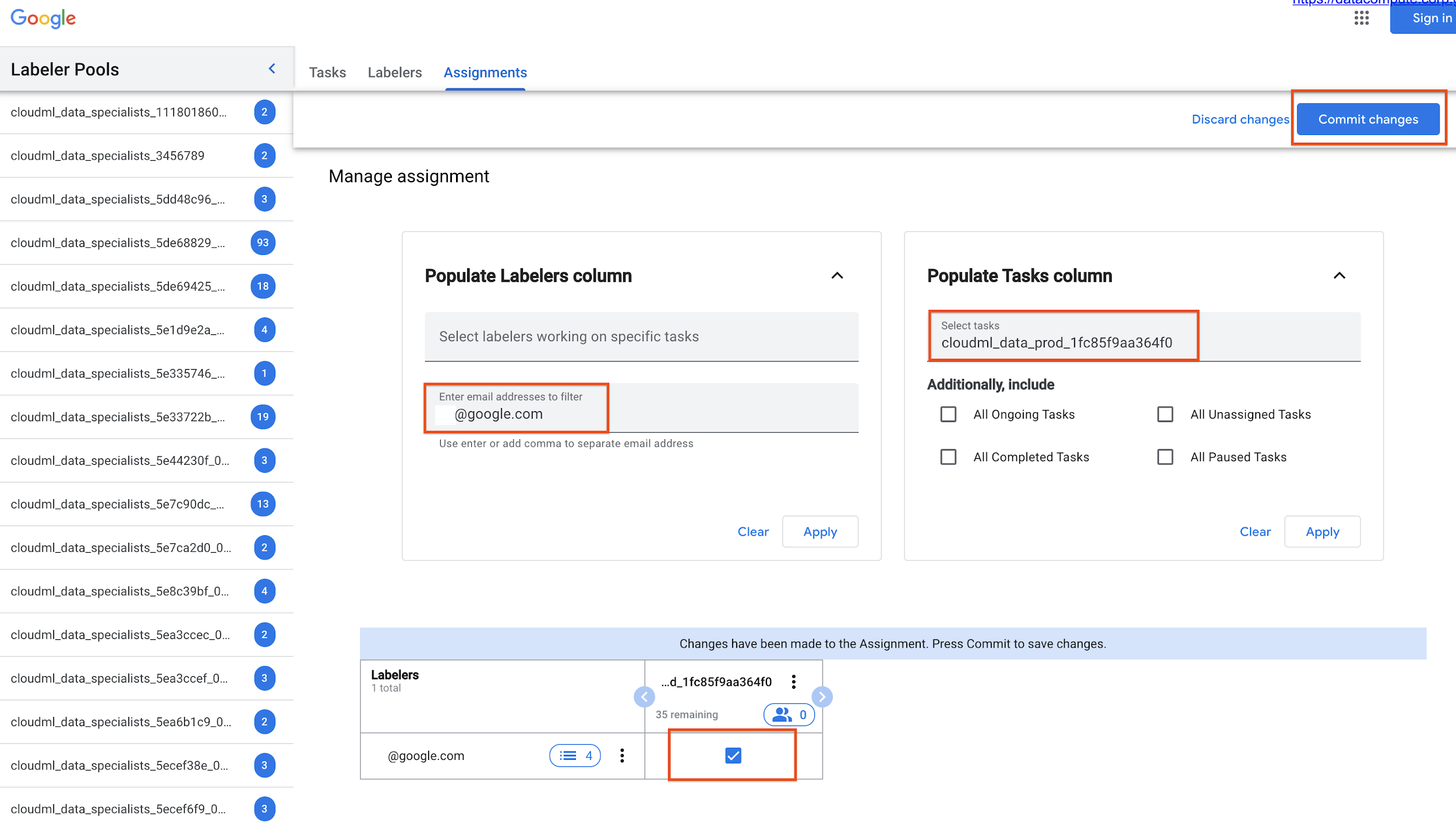
## Add Labelers

The managers need to add labelers (Google email addresses) to the labeler group in the DataCompute console. They can achieve this by following the screenshot below.



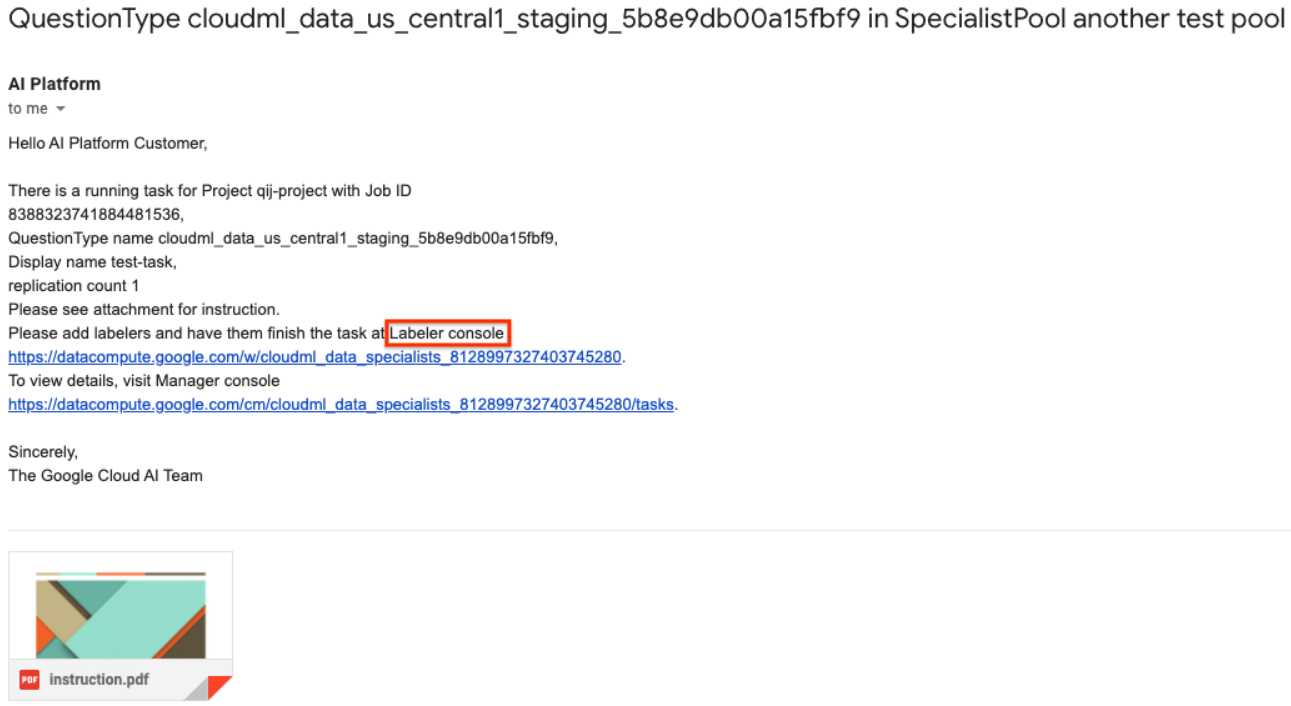
## Assign Tasks

Once labelers are added to the labeler group, labeler managers need to explicitly assign them to the labeling task by following the screenshot below

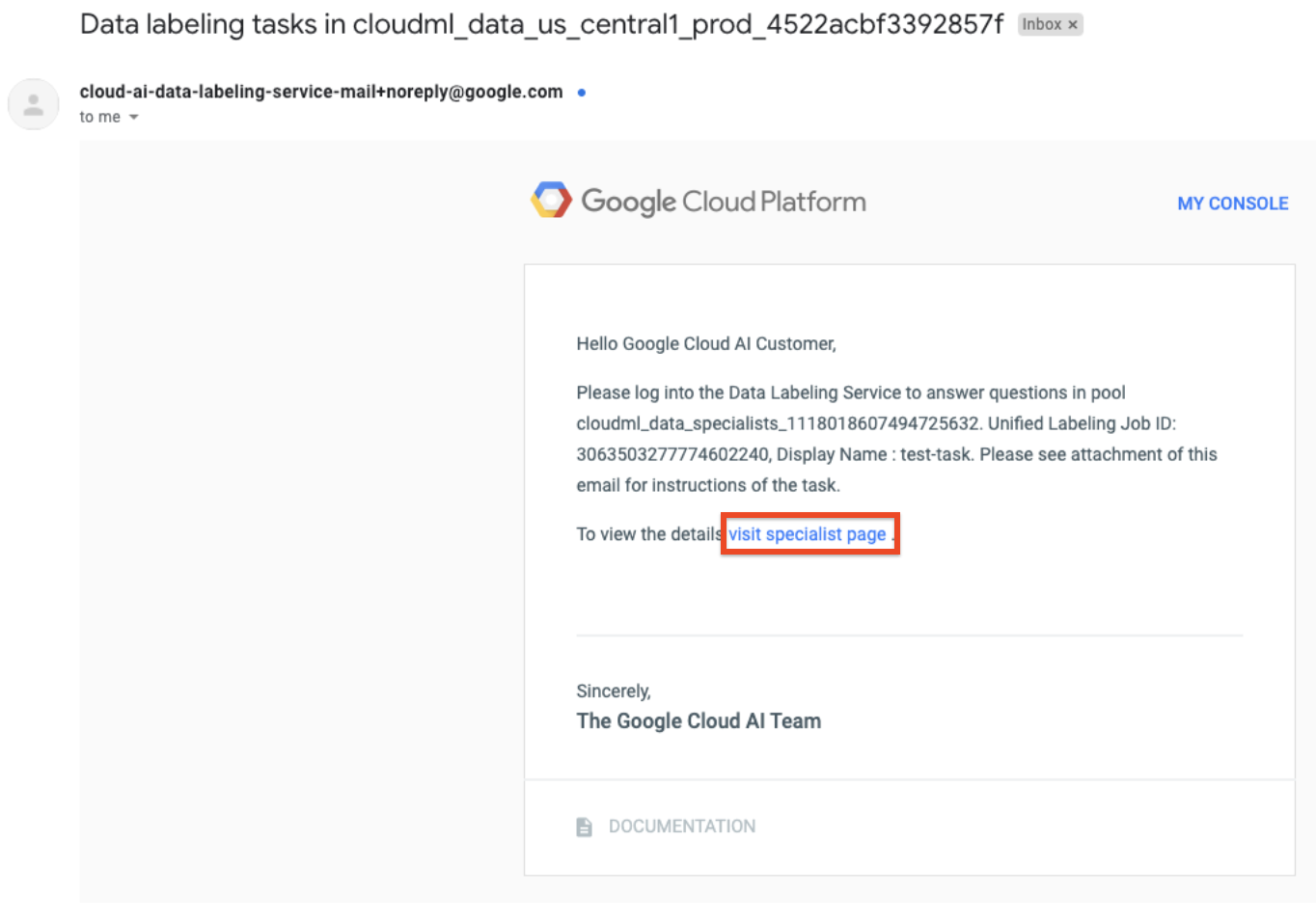


## Notify Labelers

After the assignment, labeler managers need to explicitly notify all labelers to work on the task by simply forwarding them the emails (listed below) and asking them to go to the labeler console directly[[1]](#footnote-0).



The labeler managers only need to add labelers and assign tasks to them for the first time. If the labeling task is assigned to an existing labeler group, all labelers (as well as labeler managers) in that group will receive an email (similar to the one below), which will take them to the labeling task.



# Google HITL Workforce [Private Preview]

[DOC-LINK - [https://cloud.google.com/document-ai/docs/hitl/google-workforce]](https://cloud.google.com/document-ai/docs/hitl)

Google HITL Workforce is a premium solution for customers and workloads.

This service is offered bylabelers based in India. Although the labelers are based offshore, the service offers data-at-rest residency of HITL data in either US or EU cloud regions.We are working on a near-shore [based in Poland] Workforce service, for workloads that require data-in-transit residency in the EU, but this option is currently not available.

(Note the service does not limit the locations from which Customer, Customer End Users, or Customer-selected or Google managed labeling resources may access Customer Data, or to which they may move Customer Data.)

## COVID-19 Imposed Data Restrictions

## Currently the Labeling centers are not open and the HITL workforce is working from home. As a result, they’re restricted from reviewing any documents containing PII. If your documents contain PII fields, we are restricted from offering this service until the labelers start working from the secure labeling center facilities.

## Availability

HITL Google Workforce however is in Private Preview (billable) and has limited availability.

If your document fields to be HITL-reviewed contain no PII and you would like to use Google HITL Workforce, please fill out this [Request Form](https://docs.google.com/forms/d/1l-utSiZJ51evNo6XF5li-4QR5tcmfEeRTgpIRYRUcys/edit) request access.

Pricing for HITL Platform and Workforce is provided upon providing access.

# HITL Pricing

[LINK - [https://cloud.google.com/document-ai/docs/hitl/pricing]](https://cloud.google.com/document-ai/docs/hitl)

Pricing is provided under Google NDA, and not to be shared or published externally.

## Platform Pricing

**HITL Platform:** **2c/page reviewed**, regardless of the processor used.

If you have page volumes > 500K/month, contact Sales.

## Workforce Pricing

**Pricing (Offshore labelers): 5c/field reviewed.**

* If multiple labelers are used to review the same document, you would be charged for each labeler review.
* For example, 100 documents X 5 fields per document X 3 labelers/document X 5c/field = $75
* If you have document volume > 100K/month, contact Sales.

**Managing human review costs and time** - Human validation can get expensive, whether you use your own labelers or Google labelers. The time spent per page scales with the # of fields reviewed. Google Workforce (in Private Preview) is billed per label reviewed. You can control your review effort and cost by limiting HITL to the important fields that need to be verified and corrected. To do so, use the **Label-level Validation filters** in the Human Review configuration screen.

# Labeler [Workbench](https://docs.google.com/document/d/1duvImw1GTvN7qejIYuOiSzwJoO4k5yW-HDn7IVQSx2I/edit?resourcekey=0-nMv-KgJ7L5SsSLDvfDy1wQ#heading=h.8pgfrgi1n1bk)

[LINK - [https://cloud.google.com/document-ai/docs/hitl/labeler-workbench]](https://cloud.google.com/document-ai/docs/hitl)

# Labeling [Manager Console](https://docs.google.com/document/d/1duvImw1GTvN7qejIYuOiSzwJoO4k5yW-HDn7IVQSx2I/edit?resourcekey=0-nMv-KgJ7L5SsSLDvfDy1wQ#heading=h.8pgfrgi1n1bk)

[LINK - [https://cloud.google.com/document-ai/docs/hitl/lm-console]](https://cloud.google.com/document-ai/docs/hitl)

# Creating Instructions for HITL Review

[LINK - <https://cloud.google.com/ai-platform/data-labeling/docs/instructions>]

While the HITL Labeler Workbench provides a WYSIWYG interface that maps document entities to the extracted labels and makes it easy for the labeler to compare and correct, an Instructions document is needed to instruct the human labelers what labels to look for and add, in case it’s missed by the DocAI model or Validation filters of HITL. This includes:

* Which labels to review\*
* Whether any fields are mandatory or optional\*
* Any business logic to
  + Correct labels (e.g. add “USA” for US addresses that don’t specify USA)
  + Reject documents with the correct rejection field - e.g. reject invoices >$10,000
* Special label names in the document that map to schema labels, so labeler can add these - e.g. “Client #” = “Account #”

\*These can be set up as Filters in the HITL task configuration.

HITL AI supports PDF instructions, which can provide sophisticated directions and examples for each case. You can create the instructions by creating a Google Slides and then exporting the slides as a PDF file.

## **Design good instructions**

Good instructions are the most important factor in getting good human labeling results. Since you know your use case best, you need to let the human labelers know what you want them to do. Here are some guidelines for creating good instructions:

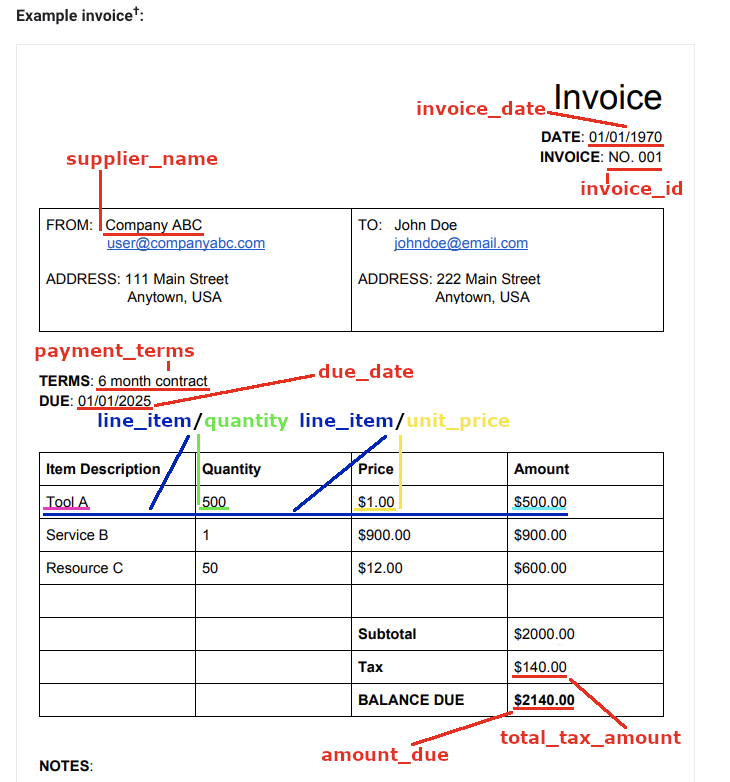
* The human labelers may not have your domain knowledge. The distinctions you ask labelers to make should be easy to understand for someone unfamiliar with your use case.
* Avoid making the instructions too long. It is best if a labeler can review and understand them within 20 minutes.
* Instructions should describe the concept of the task as well as details about how to label the data.
* If your instructions have a corresponding label set, they should cover all labels in that set. The label name in the instructions should match the name in the label set.
* It often takes several iterations to create good instructions. We recommend having a small dataset labeled first, then adjusting your instructions based on what you see in the results you get back.

A good instructions file should include the following sections:

* Label list and description: list all the labels you would like to use and describe the meaning of each label.
* Examples: For each label, give at least 3 positive examples and 1 negative example. These examples should cover different cases.
* Cover edge cases. Clarify as many edge cases as you can, This reduces the need for the labeler to interpret the label. For example, if you need to draw a bounding box for a person, it is better to clarify:
  + Do you need a box for each person if there are multiple people?
  + Do you need box if a person is occluded?
  + Do you need a box for a person who is partially shown in the image?
  + Do you need a box for a person in a picture or painting?
* Describe how to add annotations. For example:
  + For a bounding box, do you need a tight box or loose box?
  + For text entity extraction, where should the interested entity start and end?
* Clarification on labels. If two labels are similar or easy to mix up, give examples to clarify the difference.

## Visual Examples

Below is a visual example explaining to labelers where to expect different entities in the document and how they map to the extracted labels in the schema. Include such visual examples in your Instructions



1. An in-band notification mechanism is in progress. [↑](#footnote-ref-0)