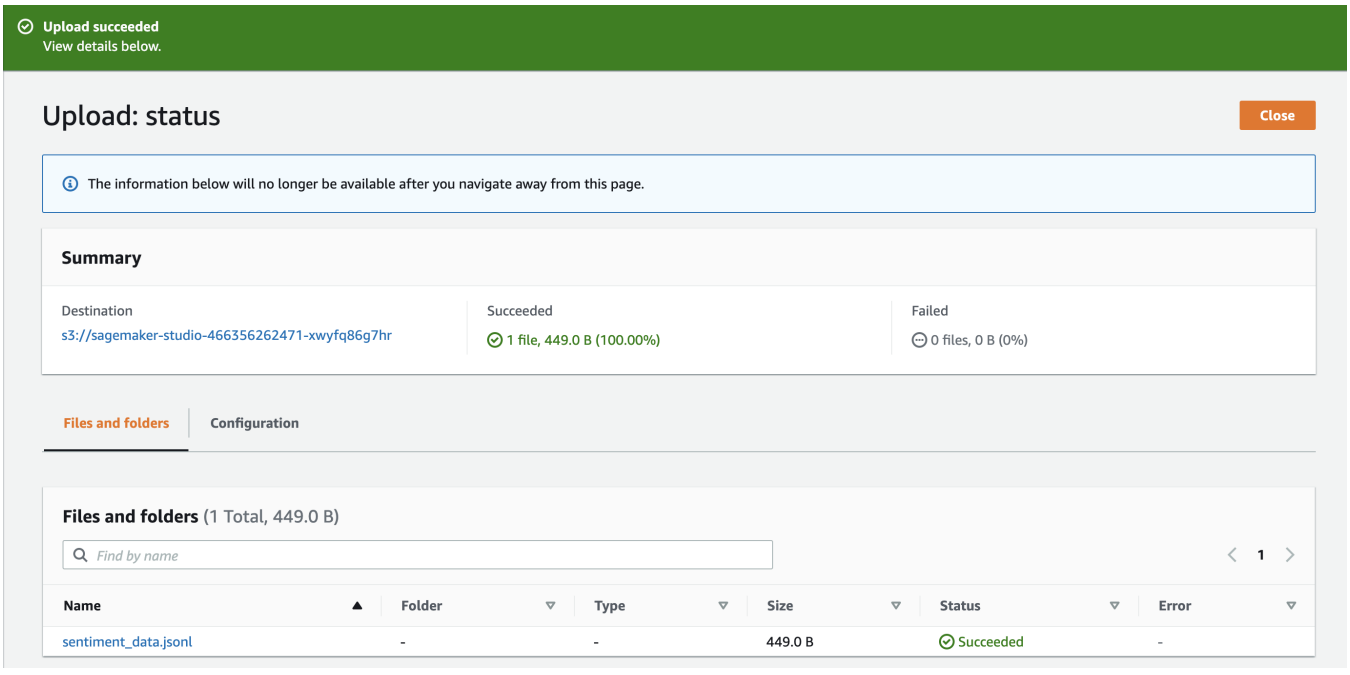
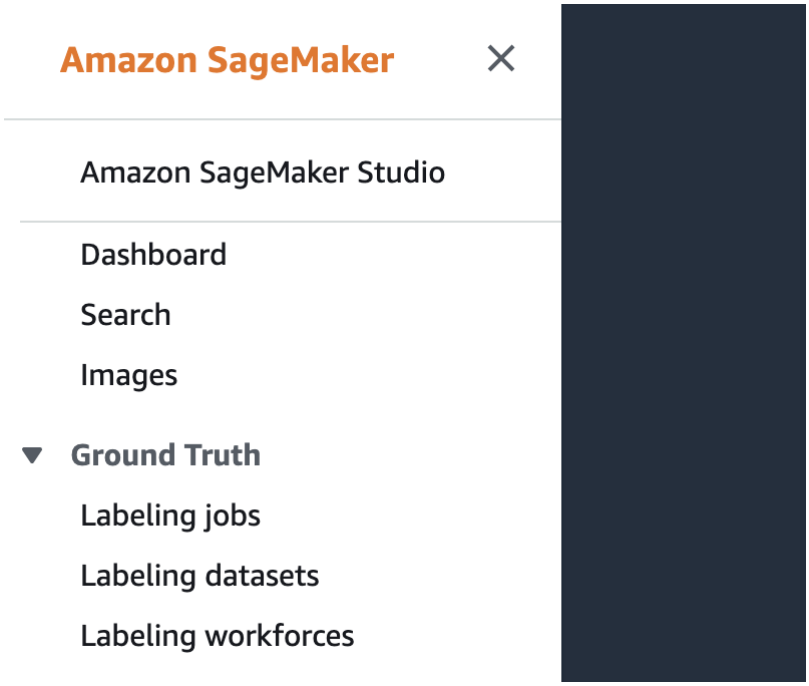
Solution: Ground Truth

Before we get started with Ground Truth, we'll need data to use in the process. In order to do that you'll need to upload the `sentiment_data.jsonl` file that is in the Github repository under this lesson's starter files. Once uploaded into a bucket, make sure you know the name of the bucket you used. We'll need it later.



S3 Data Upload

Open up the Sagemaker console in the AWS Gateway and navigate to the `Ground Truth` - `Labeling jobs` tab.

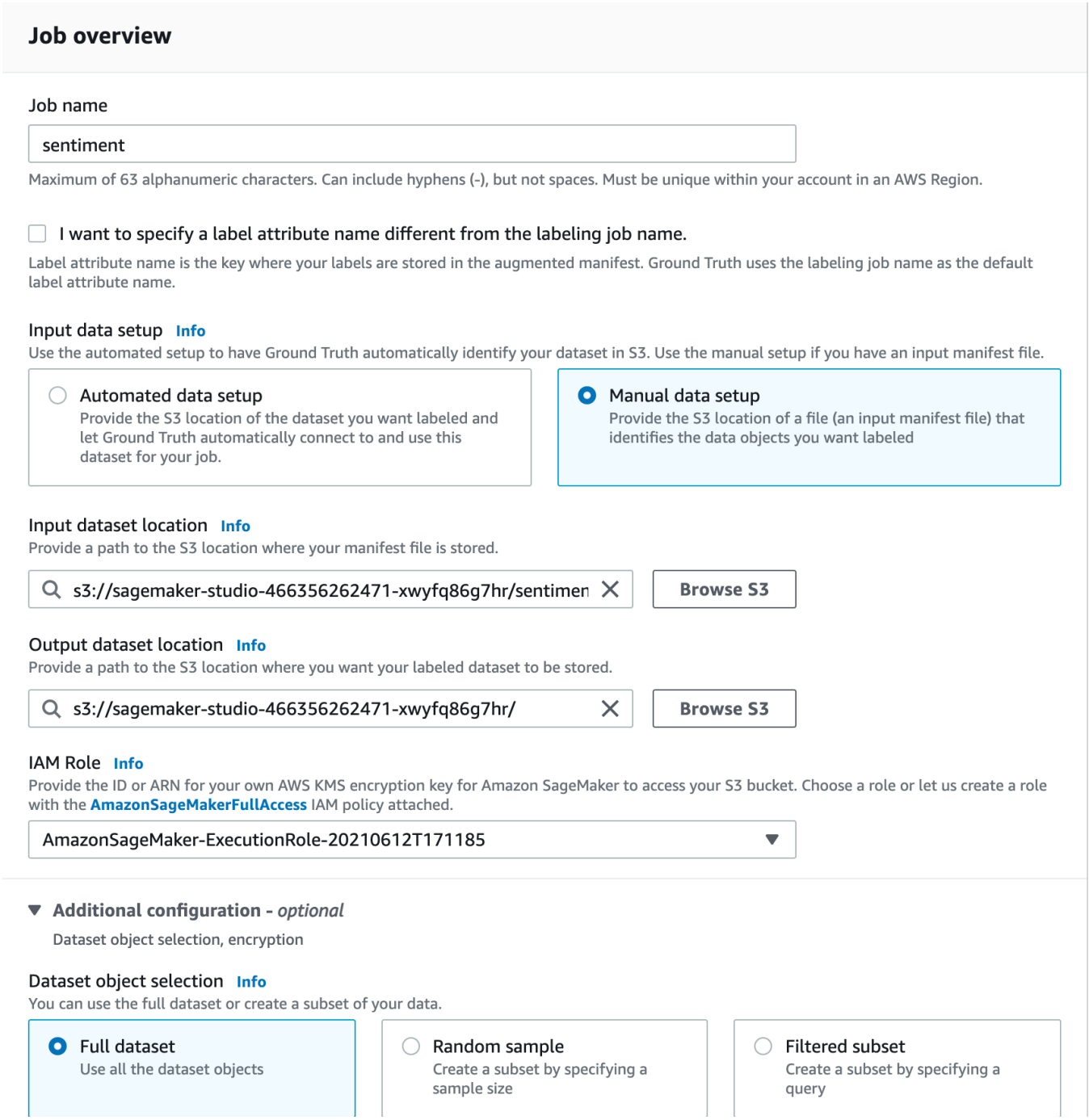


Ground Truth Page Button

When you create a labeling job, there is a bit of setup required. You'll make sure to use the `Manual data setup` option for how to input the data, as well as the S3 location of the data file you just uploaded. For the dataset output, you can just use the same S3 bucket.

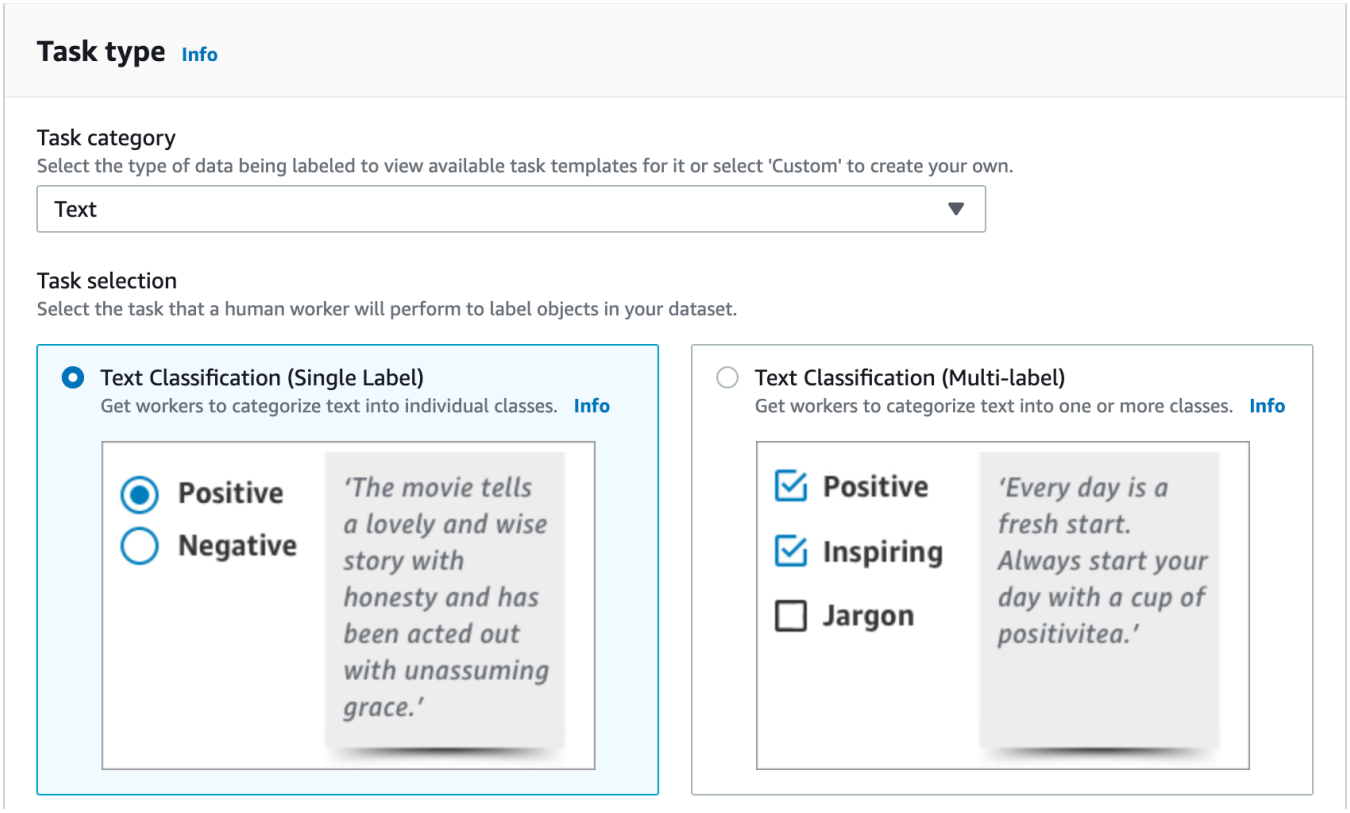
You'll need a proper Sagemaker role to execute this job. If you have not already created one from earlier in the course, you can create one now.

We'll also be using the full dataset.



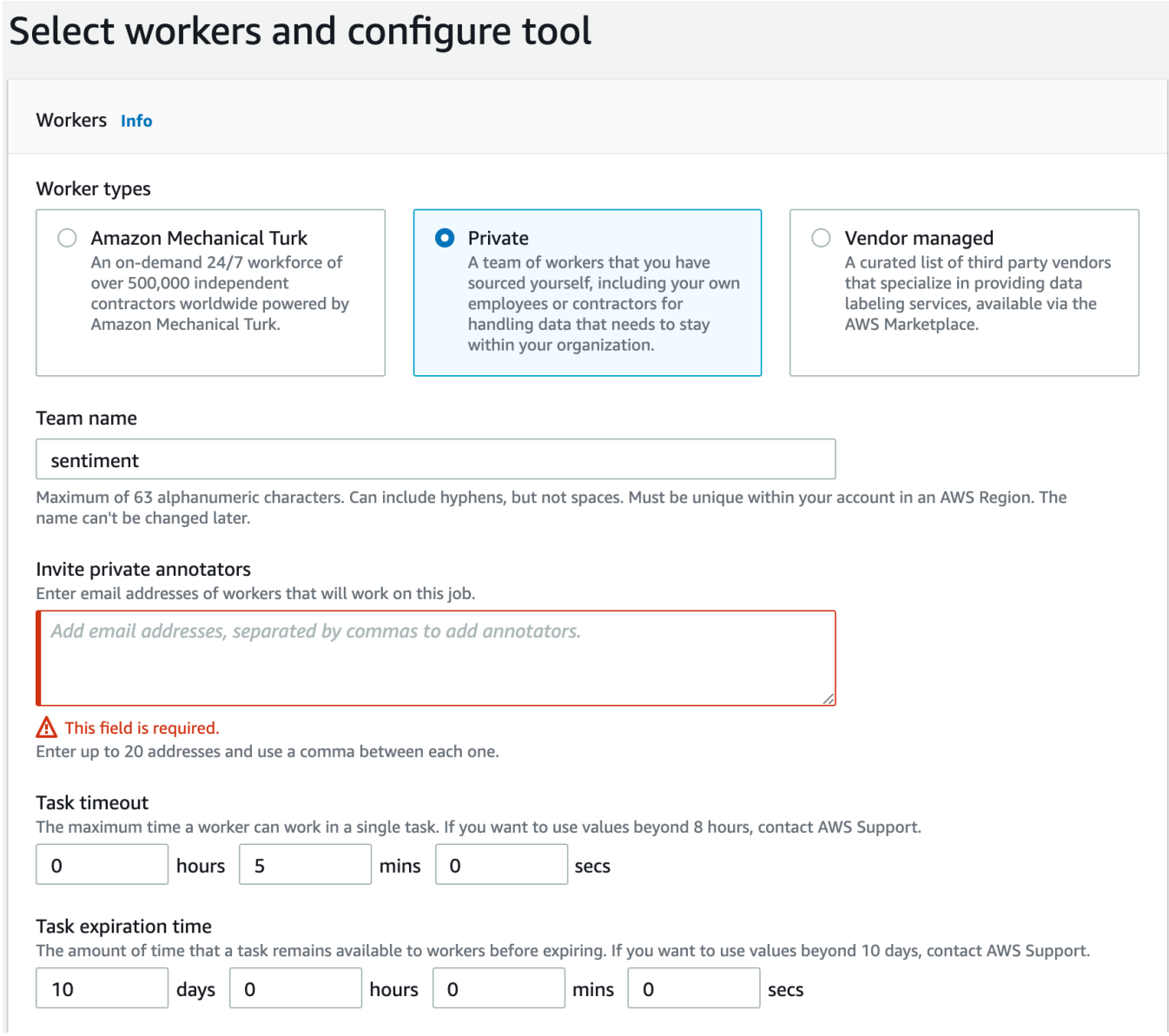
Job Overview Initial

The task type we are interested in is `Text Classification (Single Label)` . We are just interested in sentiment that is positive, negative, or neutral.



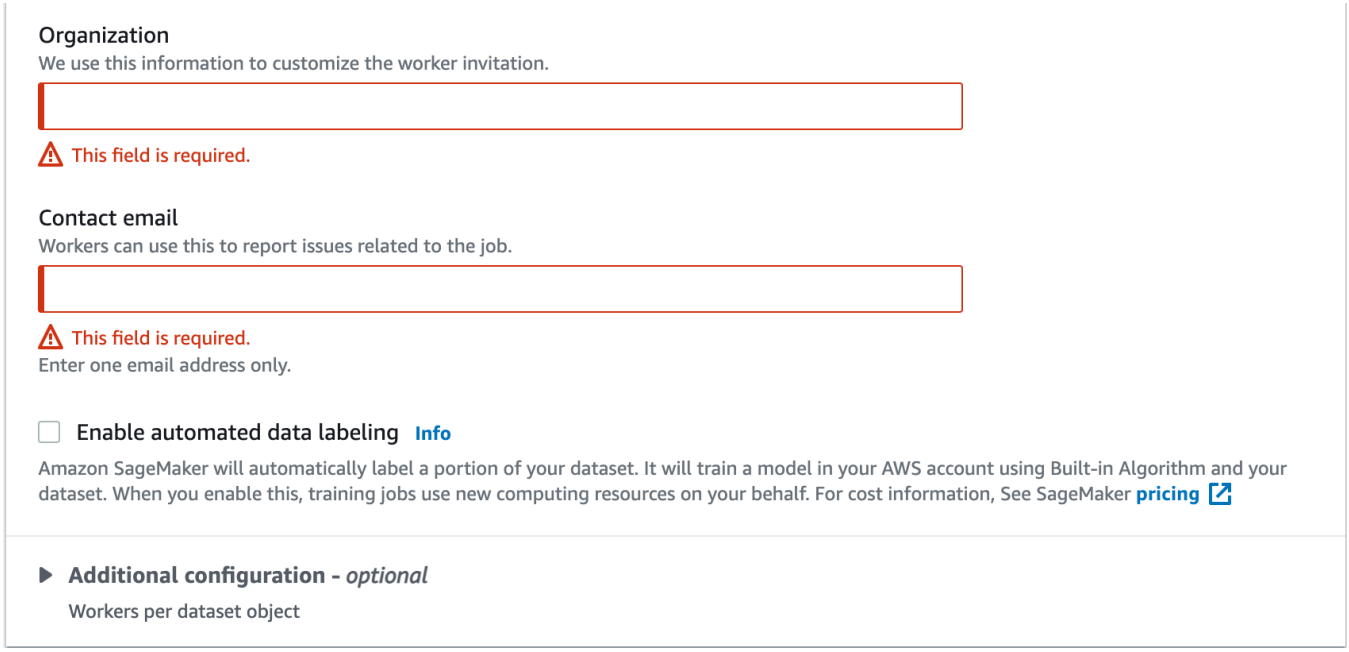
Job Overview Task

For the `Workers` section, we'll be keeping this private. More importantly, we'll just be emailing the labeling job to ourselves so we can test it out. Most of the defaults in this section are great the way they are.



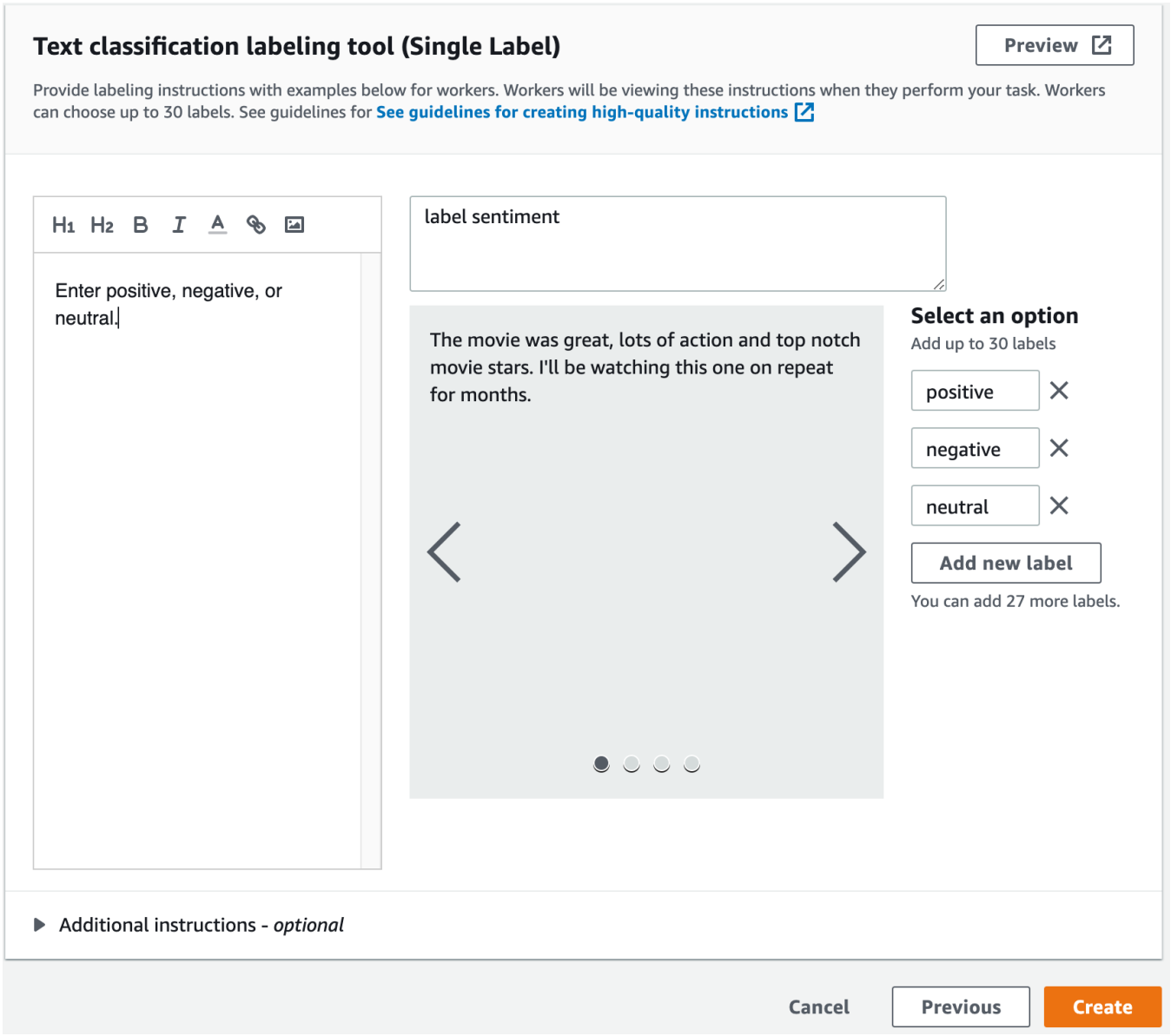
Select Private Worker

Some additional detail is needed, for `Organization` you can make up your own. Try and be creative! The `email` can be the same email you are using for annotation above.



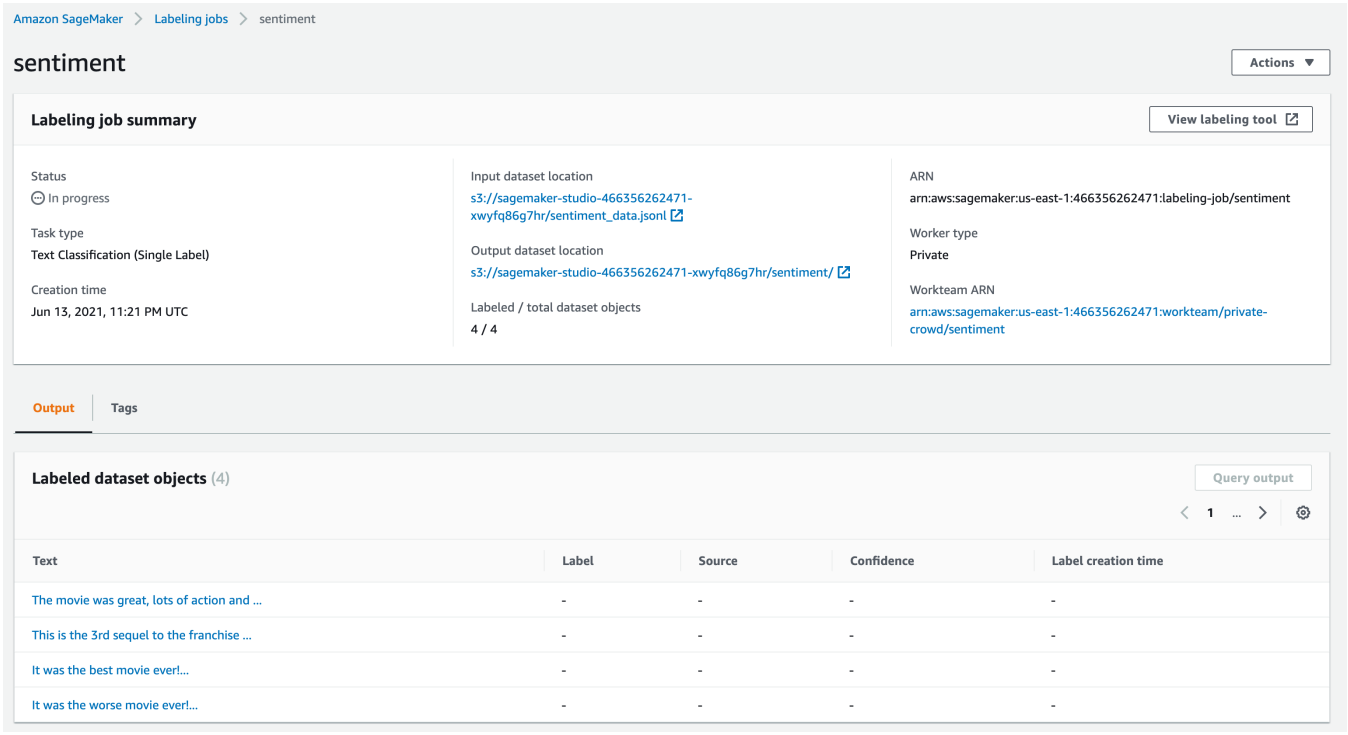
Email filling

We will define how we want to label our data in this section. The most important part is creating the `positive` , `negative` , and `neutral` labels on the right-hand side. Since we are just sending the job to ourselves, everything else in this section is optional. In is real the world, we would need to be as descriptive as possible in order to not confuse the annotator.



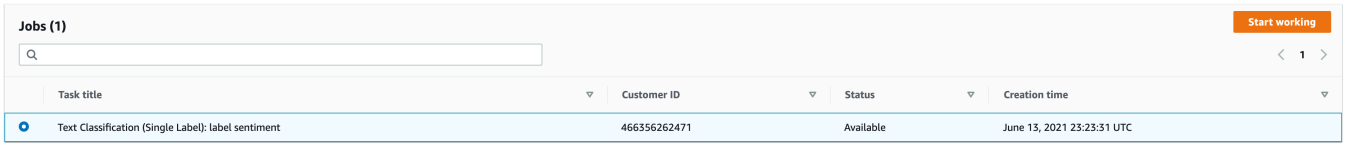
Label Options

After you create the labeling job, you will be presented with the overall summary. AWS will need to provide the necessary tooling around the annotation process, so wait a few minutes for it to set up.



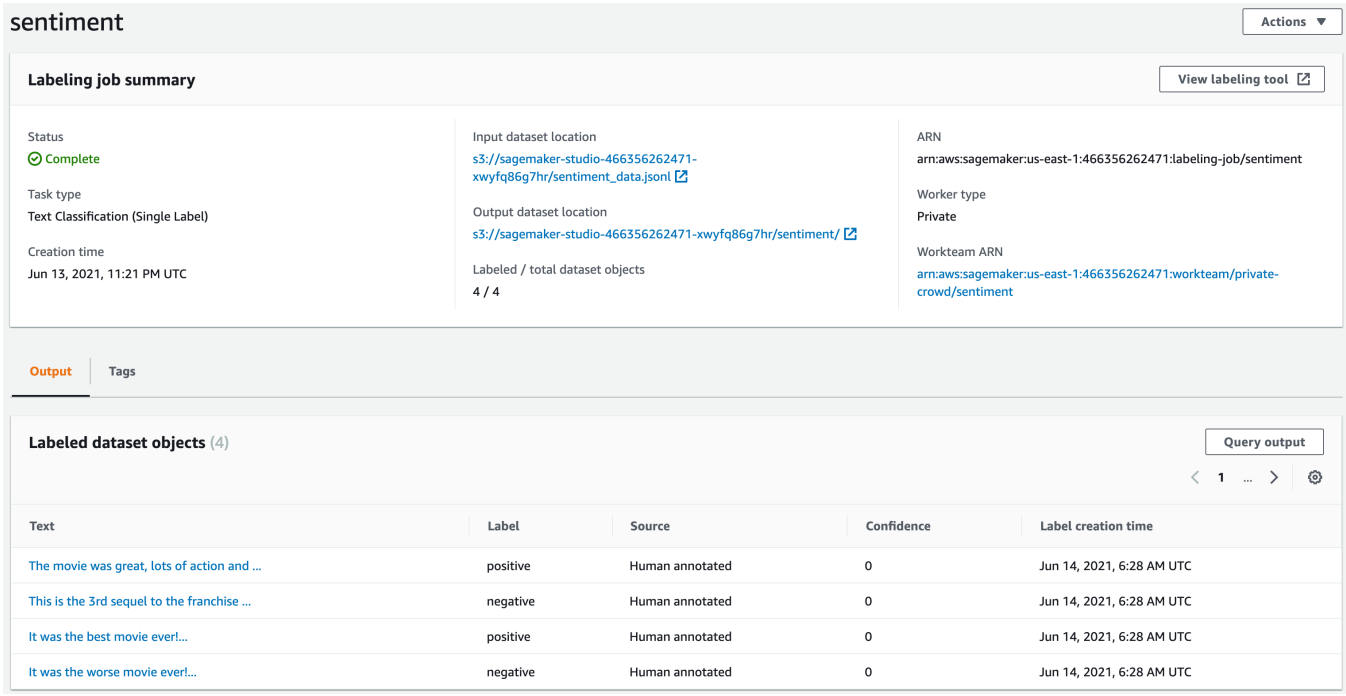
Labeling Job Summary

Shortly after creating the labeling job, you should receive an email from AWS regarding contributing to the job. You'll click through the email and use the necessary credentials to log into your Jobs panel. If you don't see your recently created labeling job, wait a few more minutes till it's ready. Eventually, you'll see something similar to below. When you do, `Start working` on the task and finish it.



Private Job

After you finished the task, you can `complete` the labeling job by stopping it from the `Actions` dropdown on the top right side of the summary page. After the job finishes, you'll be able to see results. In the `Output dataset location` navigate to S3 bucket and download the `output.manifest` file.



The output file should look like this. If it does not, make sure you received the labeling task in your email and you completed it before stopping the labeling job.

`output.manifest` File

```
{"source":"The movie was great, lots of action and top notch movie stars. I\u0027ll be watch
{"source":"This is the 3rd sequel to the franchise and you can definitely tell. It\u0027s ba
{"source":"It was the best movie ever!","sentiment":0,"sentiment-metadata":{"class-name":"po
{"source":"It was the worse movie ever!","sentiment":1,"sentiment-metadata":{"class-name":"n
```

QUIZ QUESTION