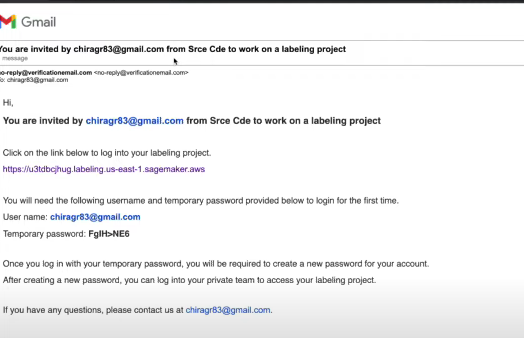
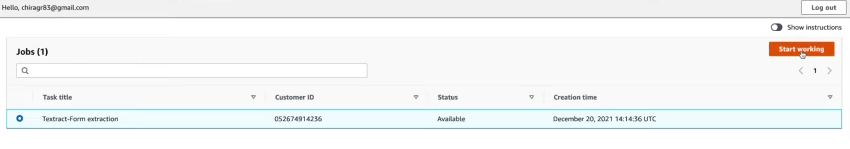
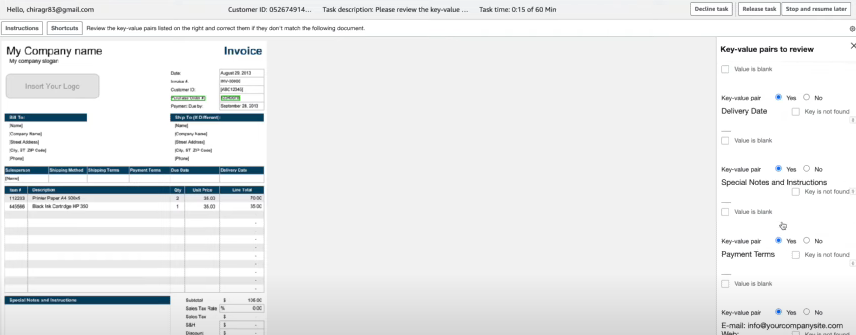
This research will look into adding a human review step for OCR. Basically, have a set of human eyes, check the parsed/scanned text after OCR has been run before the data is saved to the database. This will ensure the data's integrity and allow users to update symbols/initials for items to their full names/meanings.

<https://www.youtube.com/watch?v=JNuWFwhfO7E&list=PL5KTLzN85O4LpL7cWsFHFDsKaXtrbBgPi&index=7>

This video provides an internal human review step for OCR-scanned items. However, the issue with this implementation is that it requires the user to be invited to a different portal/site with an account to review the information read by Textract.







Although this option is definitely functional, it is very tedious for the users. This is because we will need to create an invitation to every user who has signed up to use our platform and have them switch between apps just to review whether or not the data is properly read.

An alternative solution we should do is to modify the pipeline:

User uploads file (renames and sends) → API Gateway → Lambda → S3 storage.

User Front ends up making another request to process a file with a given name → API Gateway → Lambda + Textract (on created file name) → returns JSON with formatted data to display on the user end.

At this stage, the user will see a manual input modal prepopulated with the data read by the OCR. The user will then confirm that the data is correct and the actual upload/saving of the data will be done with a previously created API.

This is a much more user-friendly design as the user would not have to hop between different applications/platforms just to scan and save his data.

Further research required:

How to use an API to save an image/document to S3 storage.