1. Document processing SaaS / software

a. SaaS solution from public cloud providers

Nowadays, many companies consider migrating their on-premises data centers to the cloud to enjoy the benefits of cloud services. Companies can pay based on the actual usage, enjoy high availability and scale with high elasticity. This can help companies reduce their maintenance cost and respond to sudden demand change quickly.

Likewise, the big three cloud providers – Amazon web services (AWS), Microsoft Azure and Google Cloud Platform (GCP) are also offering their intelligent document processing solution in software-as-a-Service (SaaS). Users upload the documents to the cloud. The SaaS solution will analyze the document layout and extract the key-value pairs based on their pre-built AI model. We can also train our custom model to achieve better performance and fit the specified use case.

b. AI-powered document processing software

Meanwhile, there are several AI-powered document processing software available in the market. This kind of software makes use of their custom AI models to analyze the document and extract the key-value pairs. Studies showed some can outperform current available AI models.

2. Robotic Process Automation (RPA)

RPA mimics human actions such as keyboard and mouse actions. With OCR tool and specified UI actions, it can analyze the key-value pair from the document with certain layouts. It is scalable and can run without human intervention (unattended RPA).

Despite being powerful, there are several concerns when using these market-available technologies as an enterprise solution.

a) Data residency and governance policy

We have to update our client's secured data to the public cloud for document processing, which may not be allowed in IT policy.

b) Black Box nature of AI models

We have no idea of the details of the AI models, including the algorithm and exact structure of their models.

c) Lack of control

We expect to have better control on the model. However, in a SaaS solution, the provider is responsible for providing the update and patch. We have to verify the correctness of the model afterwards.

So, we decided to build our AI model. In recent years, studies on machine learning have been greatly shifted to deep learning, which is a subset of machine learning using artificial neural networks such as CNN and RNN for tasks such as object detection and layout analysis. They achieve outstanding results in general.

However, there are some constraints on building these neural models - To train the model, we need to input a large (labeled) dataset with powerful GPUs and long training time.

Fortunately, many big techs have developed their open-source, transformer-based model. where we can use a smaller dataset for tuning the big, pre-trained model instead of building from scratch.

For example, YOLO for image detection. We found that there were three FYPs last year using YOLO for their object detection model. Also, DocBERT for document classification and LayoutLM and Donut for document layout analysis. These are potential models that can be employed in our project.

11/9/2023: Reduced to around 500 words.

Orange highlights are optional that can be eliminated without affecting the whole structure.

450 words if we don’t mention RPA =]