



AIM 204

# Intelligent Document Processing:

## Enhance speed and efficiency by tapping into ML

Sanket Dhurandhar  
Senior AI/ML Specialist  
AISPL

Sachin Dutta  
Chief Operating Officer  
Canara HSBC OBC Life Insurance



# Agenda for today

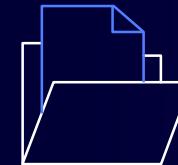
- Industry trends and challenges with document processing
- Benefits what AI/ML based Intelligent Document Processing has to offer
- The AWS Machine Learning Stack and the AI services that are key for

## Intelligent Document Processing

- Key resources to get started
- Customer Highlight : Canara HSBC OBC Life Insurance

# Documents are integral to any business

Organizations make significant operational expenditure on manually processing paper, scanned images, and electronic documents



Quotes



Forms



Invoices



Product  
Documents

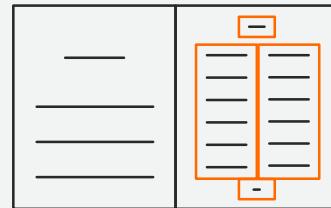


Emails



Legal  
contracts

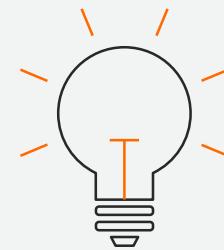
# Challenges with document extraction and analysis



Extracting text is time-consuming and introduces human error



Extracting meaningful information requires manual efforts



Building custom solutions for these tasks is expensive and hard to manage



Customizing such processes to business specific context is hard

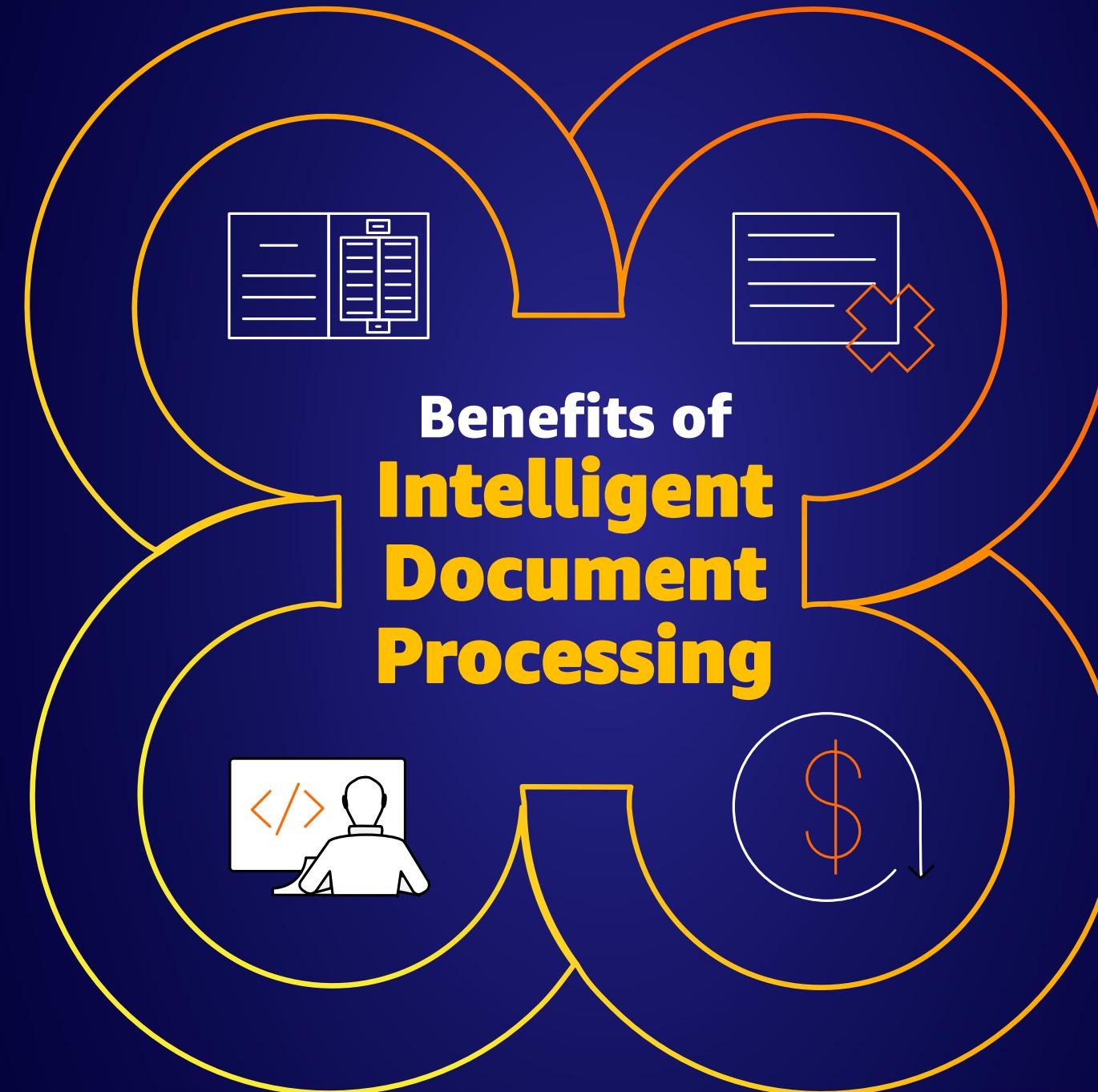
# Introducing : Intelligent Document Processing

Higher accuracy  
& Faster Data processing

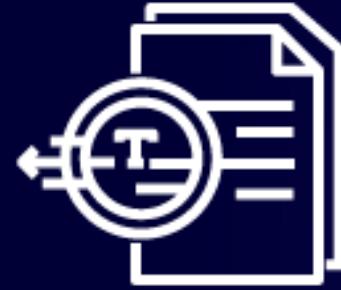
Improved employee productivity

No ML experience required

Cost Savings



# Building blocks for Intelligent Document Processing



## Amazon Textract

Quickly automate document workflows and process millions of document pages in hours



## Amazon Comprehend

Uncover insights and relationships in unstructured data



## Amazon Augmented AI

Build and manage human reviews for machine learning applications

ML-based document segmentation and data extraction, NLP-based text analysis with option to add humans in the workflow

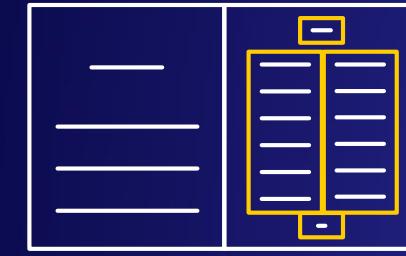
# Amazon Textract

Extract text and data from virtually any document

# Benefits of Amazon Textract



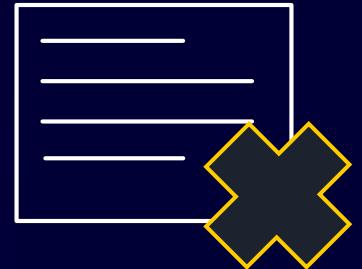
Eliminate  
manual effort  
(including template  
development)



Extract data quickly  
and accurately,  
flexible extraction  
(extract from variable  
document types)

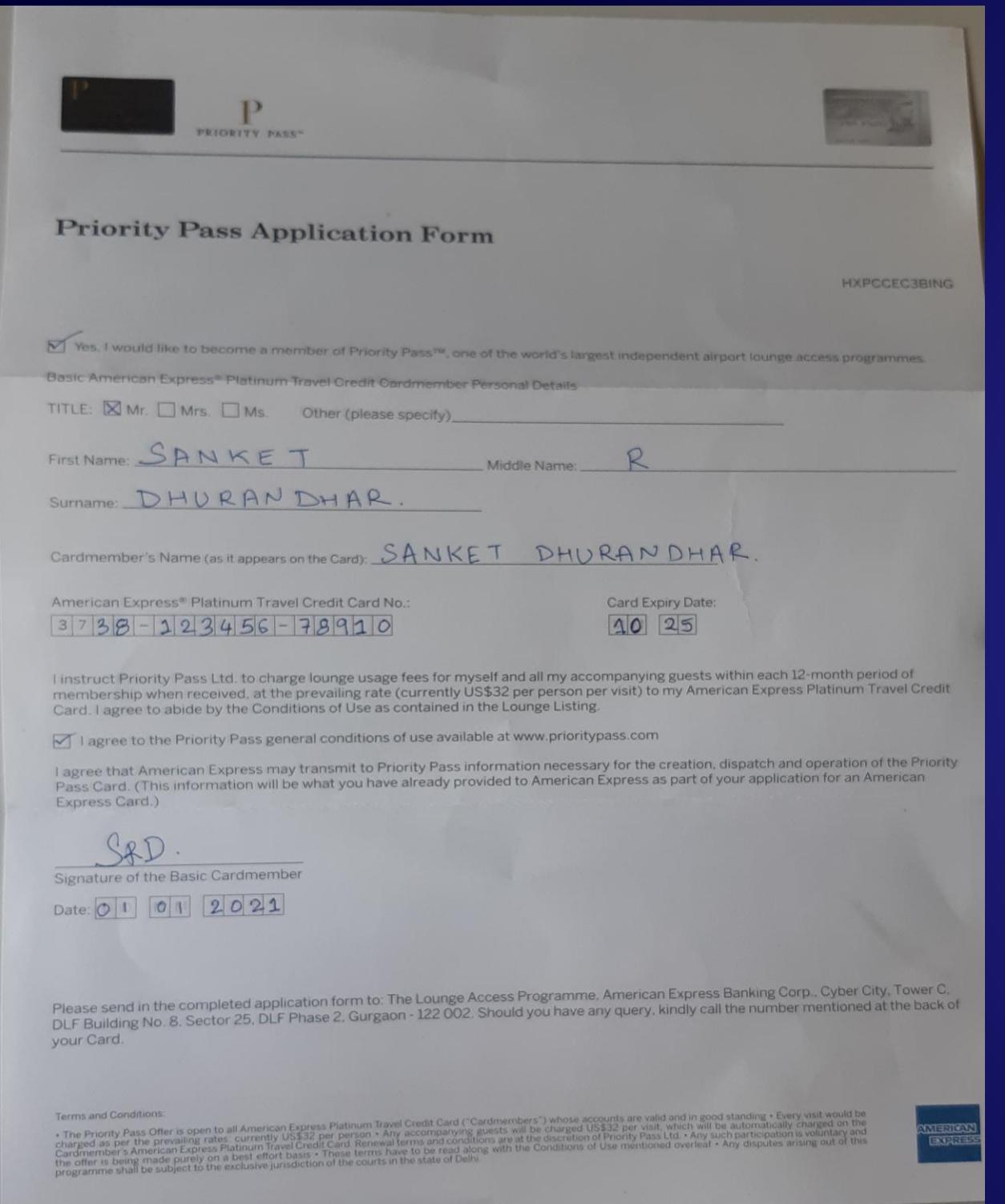


Lower document  
processing costs  
(average of 10x  
savings for OCR  
workloads)



No ML Experience  
Required

# Input form



PRIORITY PASS®

**Priority Pass Application Form**

HXPCCCEC3BING

Yes, I would like to become a member of Priority Pass™, one of the world's largest independent airport lounge access programmes.

Basic American Express® Platinum Travel Credit Cardmember Personal Details

TITLE:  Mr.  Mrs.  Ms. Other (please specify) \_\_\_\_\_

First Name: SANKET Middle Name: R

Surname: DHURANDHAR.

Cardmember's Name (as it appears on the Card): SANKET DHURANDHAR.

American Express® Platinum Travel Credit Card No.: 3738-123456-78910

Card Expiry Date: 10 25

I instruct Priority Pass Ltd. to charge lounge usage fees for myself and all my accompanying guests within each 12-month period of membership when received, at the prevailing rate (currently US\$32 per person per visit) to my American Express Platinum Travel Credit Card. I agree to abide by the Conditions of Use as contained in the Lounge Listing.

I agree to the Priority Pass general conditions of use available at [www.prioritypass.com](http://www.prioritypass.com)

I agree that American Express may transmit to Priority Pass information necessary for the creation, dispatch and operation of the Priority Pass Card. (This information will be what you have already provided to American Express as part of your application for an American Express Card.)

SRD.

Signature of the Basic Cardmember

Date: 01 01 2021

Please send in the completed application form to: The Lounge Access Programme, American Express Banking Corp., Cyber City, Tower C, DLF Building No. 8, Sector 25, DLF Phase 2, Gurgaon - 122 002. Should you have any query, kindly call the number mentioned at the back of your Card.

Terms and Conditions:

\* The Priority Pass Offer is open to all American Express Platinum Travel Credit Card ("Cardmembers") whose accounts are valid and in good standing • Every visit would be charged as per the prevailing rates, currently US\$32 per person • Any accompanying guests will be charged US\$32 per visit, which will be automatically charged on the Cardmember's American Express Platinum Travel Credit Card. Renewal terms and conditions are at the discretion of Priority Pass Ltd. • Any such participation is voluntary and the offer is being made purely on a best effort basis • These terms have to be read along with the Conditions of Use mentioned overleaf • Any disputes arising out of this programme shall be subject to the exclusive jurisdiction of the courts in the state of Delhi.

AMERICAN EXPRESS

# Console output

Cardmember's Name (as it appears on the Card):

SANKET DHURANDHAR

Middle Name:

R

American Express Platinum Travel Credit Card No.:

3738-123456-78910

Surname:

DHURANDHAR.

Card Expiry Date:

10 25

## json output

```

▼ 12 {6}
  BlockType : LINE
  Confidence : 99.14263916015625
  Text : First Name: SANKET

▼ 15 {6}
  BlockType : LINE
  Confidence : 92.38162231445312
  Text : Surname: DHURANDHAR.

▼ 16 {6}
  BlockType : LINE
  Confidence : 99.35387420654297
  Text : Cardmember's Name (as it appears on the Card): SANKET DHURANDHAR

```

## Key-Value output

key	value
Card Expiry Date:	10 25
Mr.	
Surname:	DHURANDHAR.
First Name:	SANKET
American Express Platinum Travel Credit Card No.:	3738-123456-78910
Date:	01012021
Middle Name:	R
Cardmember's Name (as it appears on the Card):	SANKET DHURANDHAR
Signature of the Basic Cardmember	SRD.

# Amazon Comprehend

A fully managed and continuously trained service that discovers insights and relationships in text

# What is Amazon Comprehend ?

Discover elements of information in text using pre-trained APIs and custom Comprehend models trained on your data.



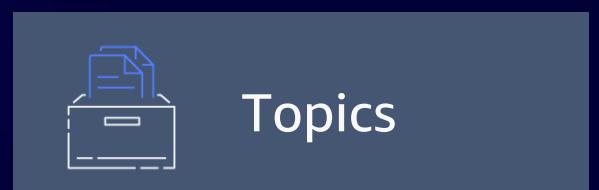
Documents

Email, chat,  
social, phone  
calls and more



Amazon  
Comprehend

Automatically  
extract insights  
from text



✓ Short time-to-value

✓ Low learning ramp

✓ Always improving accuracy

# Amazon Comprehend learns your organization's language

## Annotated Training Examples

Example				
File	Line	Begin Offset	End Offset	Type
documents.txt	0	0	12	ENGINEER
documents.txt	1	0	5	ENGINEER
documents.txt	3	25	30	MANAGER



Documents containing terms and phrases in text

## Comprehend Training



- Automated annotation
- Automated algorithm selection
- Automated tuning and testing
- SDK or code-free Console UX

## Custom Classifiers



## Custom Entity Recognizers



# Amazon Augmented AI (A2I)

Easily implement human review of machine learning predictions

# Benefits with Amazon A2I



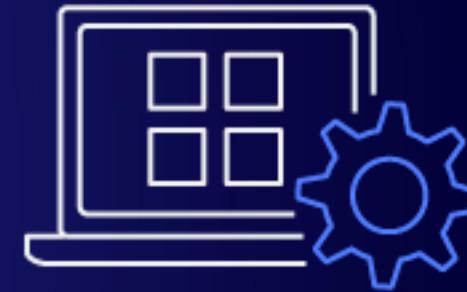
**Easily implement  
human review  
workflows**



**Reduce time to market  
with pre-built  
workflows  
and UIs**



**Multiple workforce  
options**



**Integrate with  
your custom ML  
models**



**Increased worker  
efficiency**

# Human Review Workforces



## Amazon Mechanical Turk

An on-demand 24x7 workforce of over 500,000 independent contractors worldwide, powered by Amazon Mechanical Turk



## Private

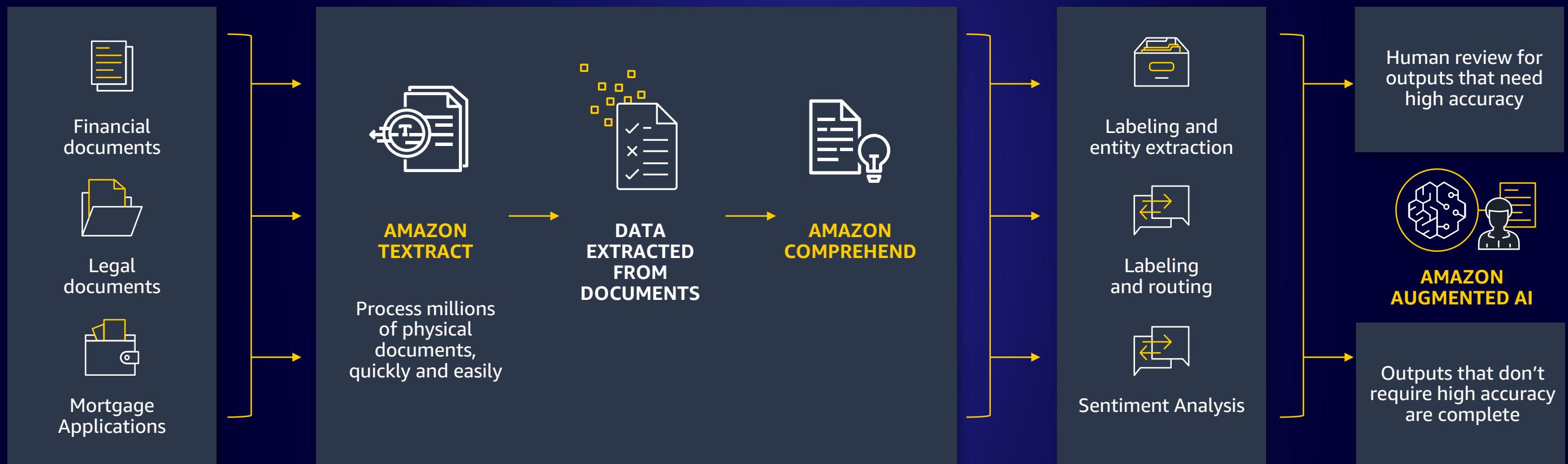
A team of workers that you have sourced yourself (your own employees or contractors)



## Vendors

AWS Marketplace vendors that specialize in human review tasks

# Bringing it all **together**



# Key resources to get started

## Intelligent Document Processing

<https://aws.amazon.com/machine-learning/ml-use-cases/document-processing/>

## Amazon Comprehend

<https://aws.amazon.com/comprehend/>

## Amazon Textract

<https://aws.amazon.com/textract/>

## Amazon Augmented AI (Amazon A2I)

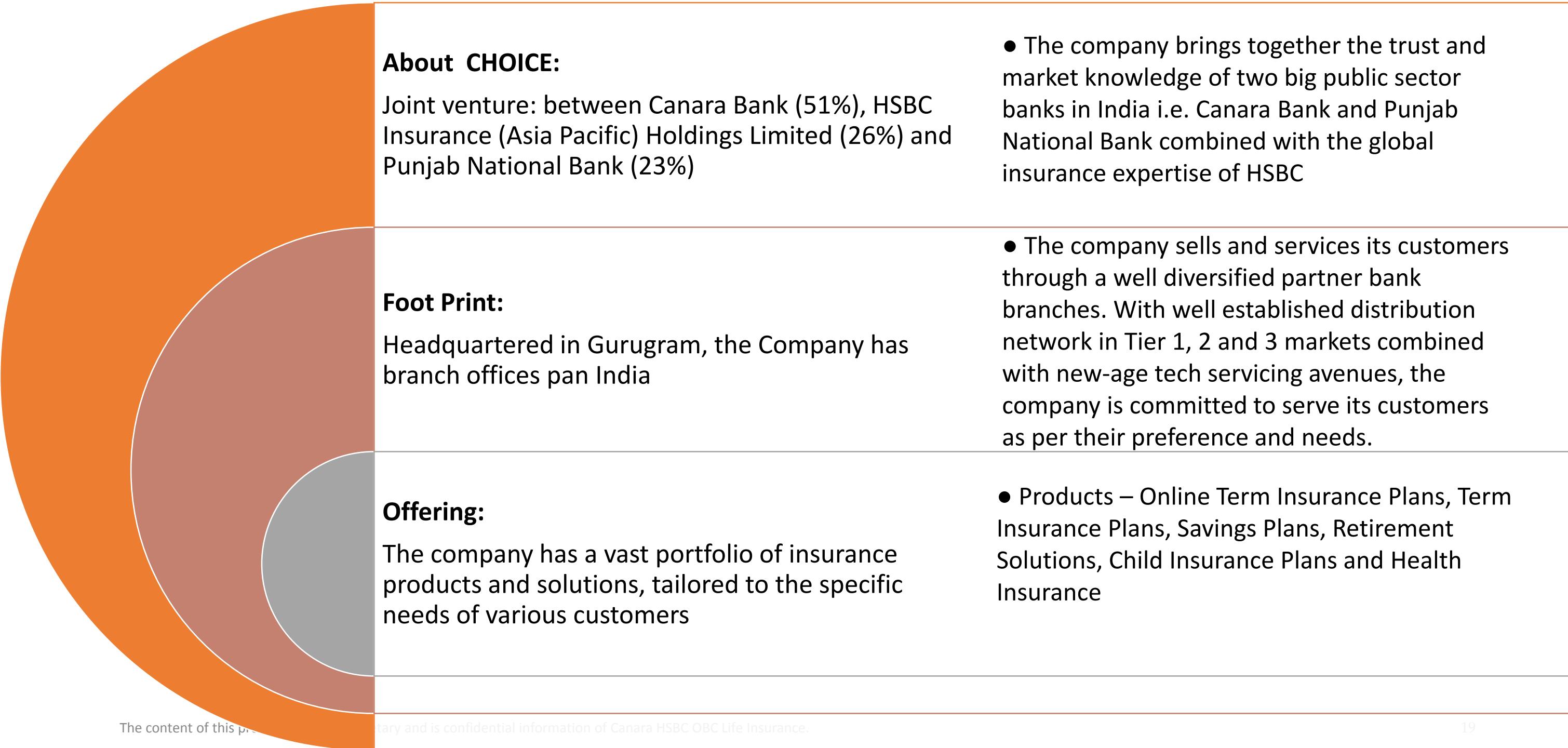
<https://aws.amazon.com/augmented-ai/>

## Document Analysis Solution – GitHub Sample

<https://github.com/aws-samples/amazon-textract-comprehend-a2i>

**Significantly  
enhancing speed  
and efficiency  
by harnessing machine  
learning capabilities**

# Canara HSBC Oriental Bank of Commerce Life Insurance (CHOICE)



# Business Need, Finalizing Amazon Textract, Experience and Benefits

We had a manual activity being carried out with --- resources carrying out the activity for past and new data extraction.  
Past data review was to be completed in 8-9 month and was prone to manual failure

Scope: Automation of Document reviews and Other Requirements

- Textract Delivered Highest Accuracy
- Easy to Use and Align with Business Solution/ Workflows

POC

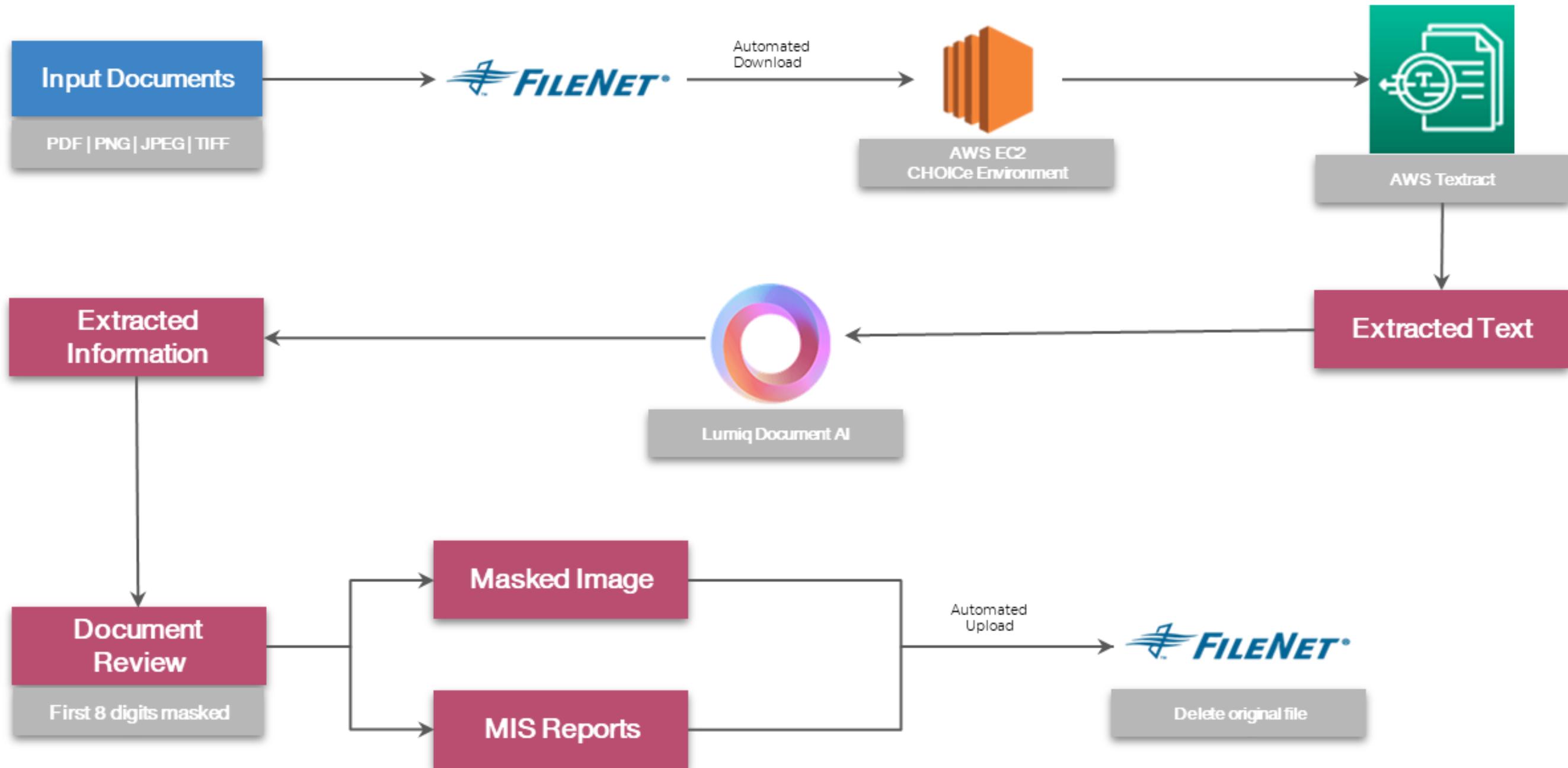
Implementation

- Quick
- Up and Running within weeks
- Developer managing custom actions

- 92% during Initial Run
- Increased to 97% after Multiple Runs
- Completed the activity in 3 weeks

Accuracy Achieved

# Solution Architecture



# ● Future Plans of Intelligent Document Management

## New Business

Complementary Products

Demographics, Sales, Product,  
Financial Information

## Fraud Detection

Non Disclosure of Health  
Information

Non Traceable Customer  
& Death Before Issuance

- Less conversion considering large customer base
- Lack of data driven product pitching
- Product Categories are not well known
- Different marketing and sales channels

- Huge number of Logins and Issuance
- High Investigation/ Background Verification Cost
- Spike in Frauds percentage

Created a small team with vendor support to automate operational tasks which provides higher confidence on processing with high accuracy and confidence score allowing completely System based process



# Thank you!

Sanket R. Dhurandhar

India Head – AI & Machine  
Learning, AISPL

Sachin Dutta

Chief Operating Officer  
Canara HSBC OBC Life Insurance





Please complete  
the session survey

