



# Modernize your document workflows

Intelligent Document Processing

Brijesh Pati

Enterprise Solutions Architect

# Agenda



## **Document Processing Challenges**



## **AI/ML overview**



## **Amazon Intelligent Document Processing (IDP)**



## **Demo**



# Legacy document processes do not meet today's needs



Legacy OCR and manual processes are time-consuming, error-prone, and expensive



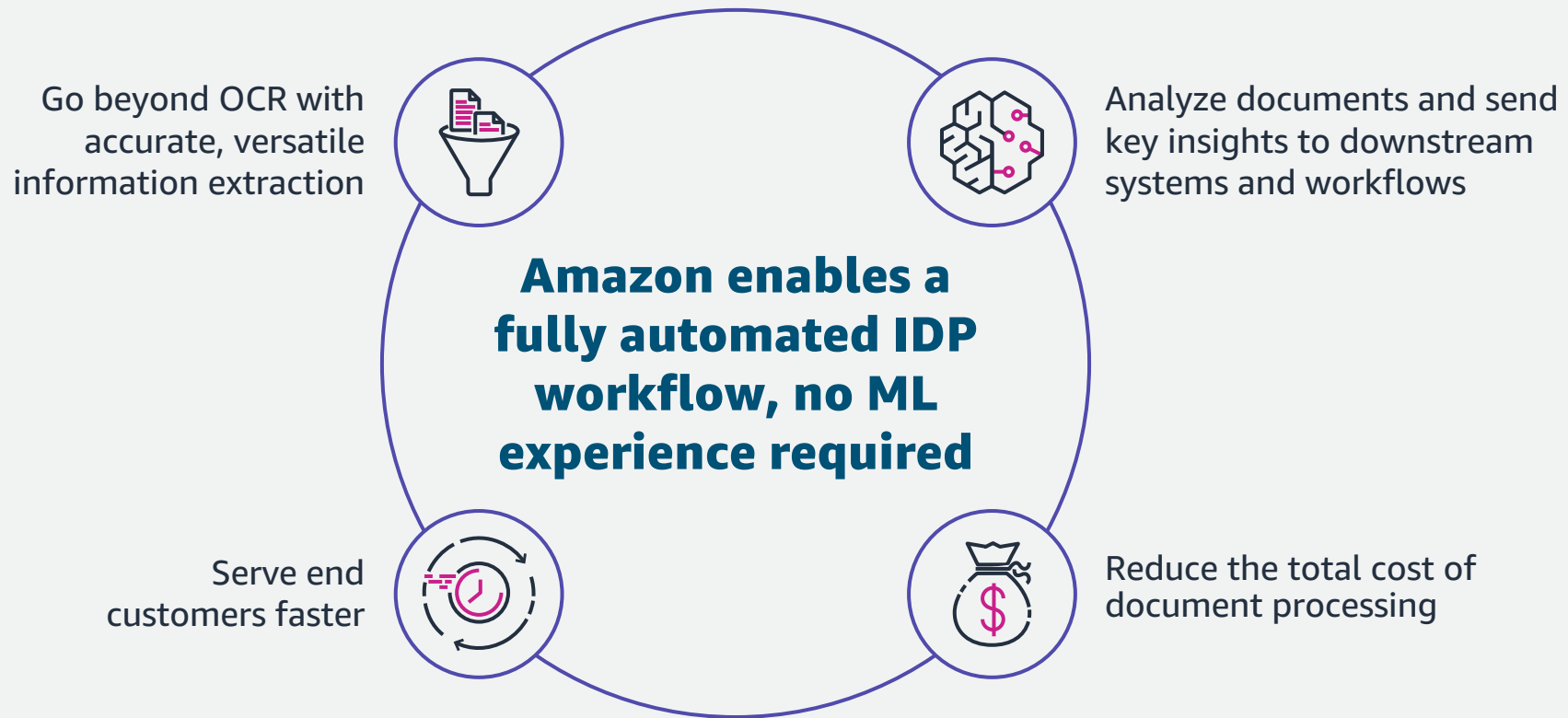
Manual processes do not scale easily with document volume



Difficult to find useful information needed for business decisions

**Faster document processing shortens decision cycles;  
this enables you to serve more customers and have people do higher value tasks**

# Benefits of Amazon's Intelligent Document Processing (IDP)



# Analyze documents and send key insights to downstream systems and workflows

## Classify type of document



- NDA
- Purchase agreement
- Invoice

## Find Key Data

CO: 10543 DEPT: 12345 NUMBER: 1  
ABC 10543 12345 12345 00000000  
ANY COMPANY CORP  
456 ANY AVENUE  
ANYTOWN, USA 12345

Serial Security Number: 987654321  
Employee Name: John Stiles  
Employment Information:  
Federal: 3-501 Additional Tax  
State: 2

Earnings

	rate	hours	this period	year to date
Regular	10.00	32.00	320.00	16,640.00
Overtime	15.00	1.00	15.00	760.00
Holiday	10.00	8.00	80.00	4,160.00
Tuition			37.47	1,646.80
			<b>\$ 452.47</b>	<b>23,526.80</b>

Deductions

	rate	hours	this period	year to date
Federal Income Tax	-40.00		2,111.00	
Social Security Tax	-24.00		1,456.80	
Medicare Tax	-6.00		341.12	
NY State Income Tax	-6.00		438.36	
NYC Income Tax	-5.00		308.48	
NY SUTS/CA Tax	-0.00		31.20	
Other	-5.00		100.00	
401(k)	-28.00		1,500.00	
Sick Plan	-15.00		150.00	
Life Insurance	-5.00		50.00	
Loan	-30.00		150.00	
Adjustment			13.50	
UTA Insurance				
<b>Net Pay</b>			<b>\$ 395.90</b>	

\* Excluded from federal taxable wages  
Your federal wages this period are \$386.15

Earnings Statement

Period ending: 7/18/2008  
Pay date: 7/29/2008

JOHN STILES  
101 MAIN STREET  
ANYTOWN, USA 12345

Other Benefits and Information

	this period	total to date
Group Term Life	0.71	27.00
Loan Amt Paid		840.00
Vac Hrs		40.00
Sick Hrs		10.00
Time		Operator

Important Notes

EFFECTIVE THIS PAY PERIOD YOUR REGULAR HOURLY RATE HAS BEEN CHANGED FROM \$8.00 TO \$9.00 PER HOUR.

WE WILL BE STARTING OUR UNITED WAY FUND DRIVE SOON AND LOOK FORWARD TO YOUR PARTICIPATION.

## Output

Question: What is the Gross Pay YTD?  
Answer: \$23,526.80  
Question: What is the employee address?  
Answer: 101 Main Street Anytown, USA 12345

## Recognize entities specific to your business

### Document text

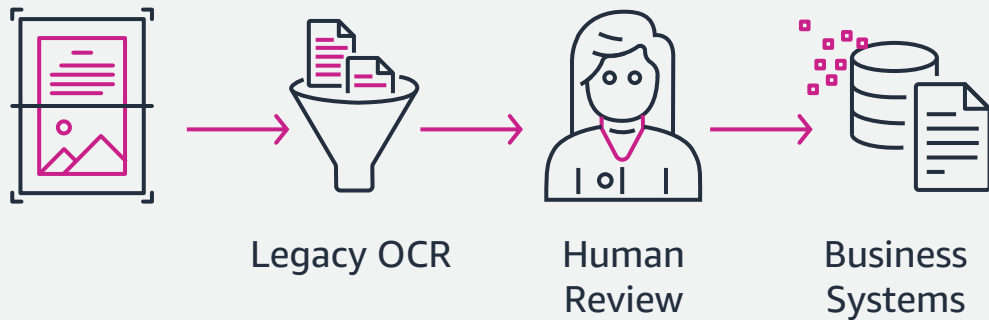
Hi my name is John. I am filing a claim for a car accident, my insurance number is 456-YQT.

## Output

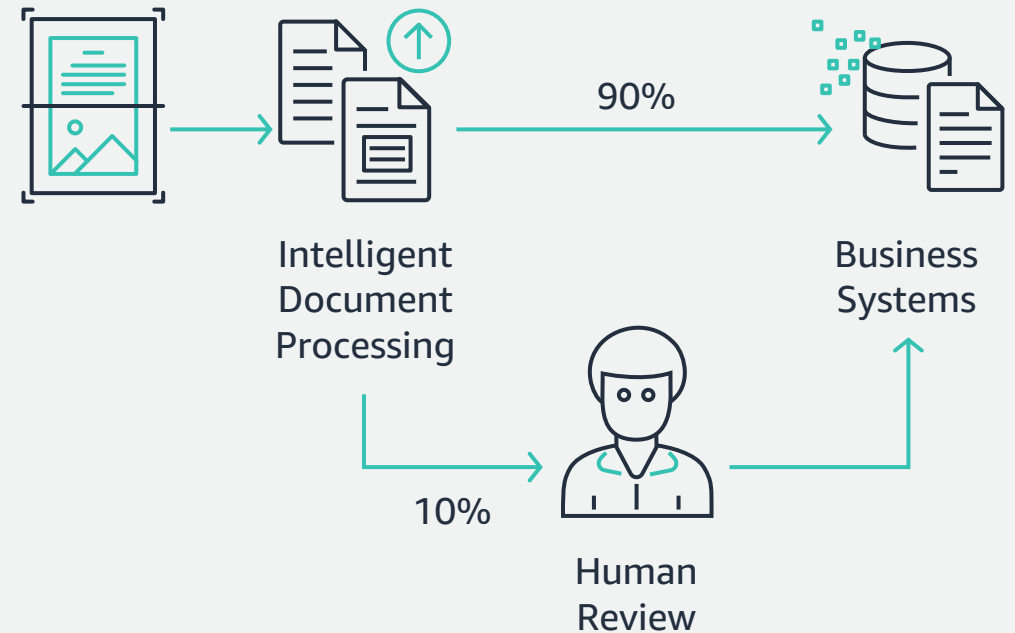
Category of Entity: Policy\_ID  
Entity: 456-YQT  
Confidence: 0.95

# Reduce the total cost of your document processing

Common for people to review all documents prior to downstream usage

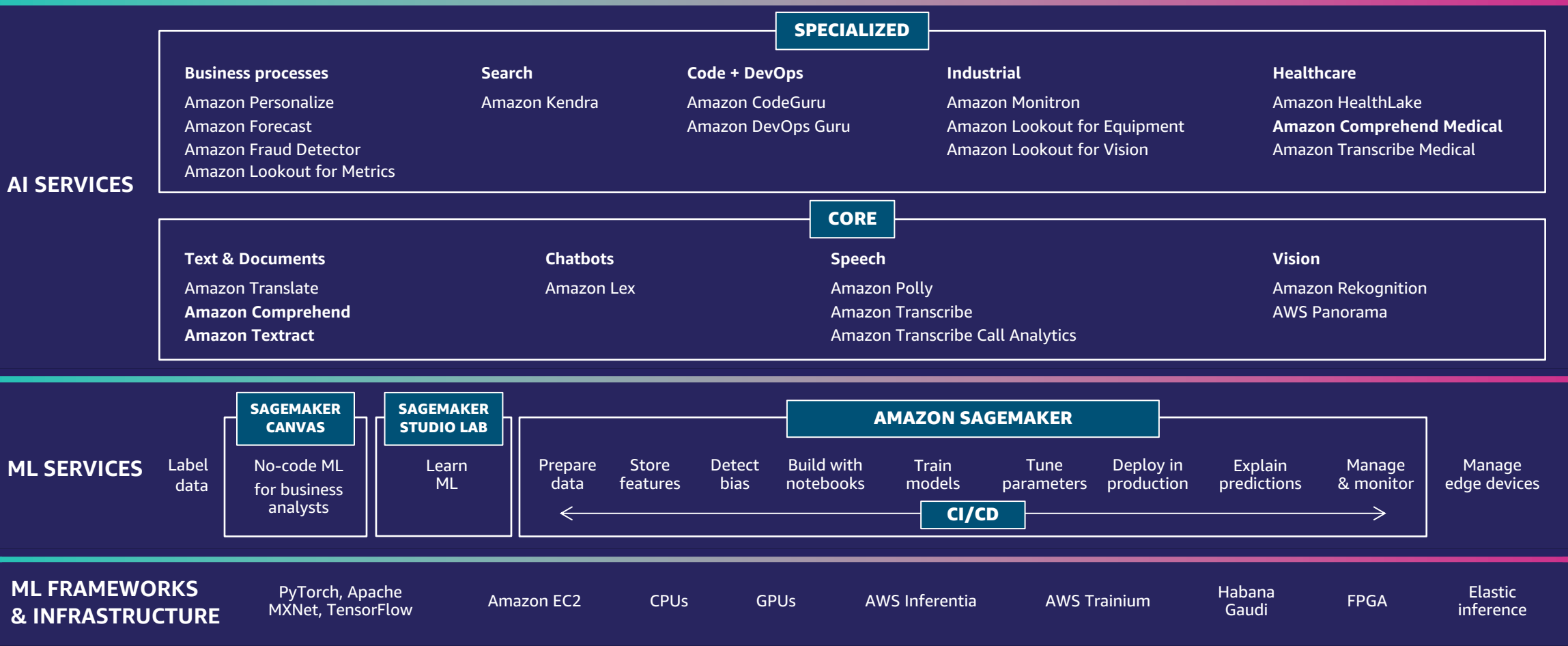


Limit the people who review the documents with confidence scores < target



# The AWS AI/ML stack

## Broadest and most complete set of machine learning capabilities





# Building / deploying your own model vs. AI solutions comparison

	<b>SageMaker or Your Own ML Model</b> Prepare, build, train, and deploy high-quality machine learning models	<b>AI you can customize</b> <b>Comprehend</b> Custom Classifier/ Custom Entity Recognition	<b>Pretrained AI services</b> Textract, Comprehend, and Comprehend Medical
<b>ML development experience</b>	✔ Required	Not required	Not required
<b>Data preparation</b> Data collection, prep, transformation	✔ Required	✔ Limited effort required	Not required
<b>Model selection and training</b> Select and appropriate algorithm to train a model	✔ Required	Not required	Not required
<b>Model tuning</b>	✔ Required	✔ Limited effort required	Not required
<b>Model deployment, monitoring, and maintenance</b>	✔ Required	Not required	Not required
<b>Easy to use API for real-time and batch inferences</b>	Must be implemented	Ready to use after limited training data provided	Ready to use



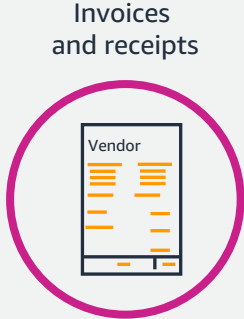
# Amazon Textract - Features



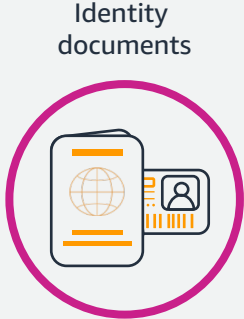
**Text**



**Forms**



Invoices  
and receipts



Identity  
documents

**Specialized documents**



**Handwriting**



**Tables**

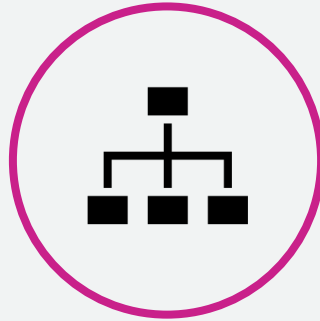


**Queries**

# Amazon Comprehend - Features



**Entity Recognition**



**Classification**



**Sentiment Analysis**



**PII Identification  
and Redaction**



**Key Phrases**



**Syntax Analysis**

MORTGAGE STATEMENT

Statement Date: MM/DD/YYYY

1-111-TJNUM\_1234567-111-2-333-444-555-666

Customer Name 1

Customer Name 2

Address 1

Address 2

City State ZIP

2

Account Number: 9999999999

Payment Due Date: 4/1/2012

**Amount Due:** \$1,043.00

*If payment is received after MM/DD/YYYY a \$37.46 late fee will be charged.*

Property Address:

Address 1

Address 2

City, State Zip Code

3

Account Information

Outstanding Principal Balance	\$151,243.38
Deferred Principal Balance	\$0.00
Escrow Account Balance	\$0.00
Unapplied Funds Balance	\$0.00
Interest Rate (until mm/dd/yyyy)	4.25%
Loan Maturity Date	9/1/2032
Prepayment Penalty	None
Taxes Paid Year-to-Date	\$0.00

5

Past Payments Breakdown

	Paid Last Month	Paid Year to Date
Principal	\$212.82	\$2,090.00
Interest	\$536.41	\$3,167.03
Escrow	\$293.77	\$2,937.70
Fees	\$0.00	\$0.00
Other	\$0.00	\$0.00
Unapplied Funds	\$0.00	\$0.00
Total	\$1,043.00	\$8,194.73

Explanation of Amount Due

Principal	\$749.23
Interest	\$0.00
Escrow (for Taxes and/or Insurance)	\$293.77
Other	\$0.00
<b>Regular Monthly Payment</b>	<b>\$1,043.00</b>
Fees and Other Charges	\$0.00
Overdue Payment	\$0.00
<b>Total Amount Due</b>	<b>\$1,043.00</b>

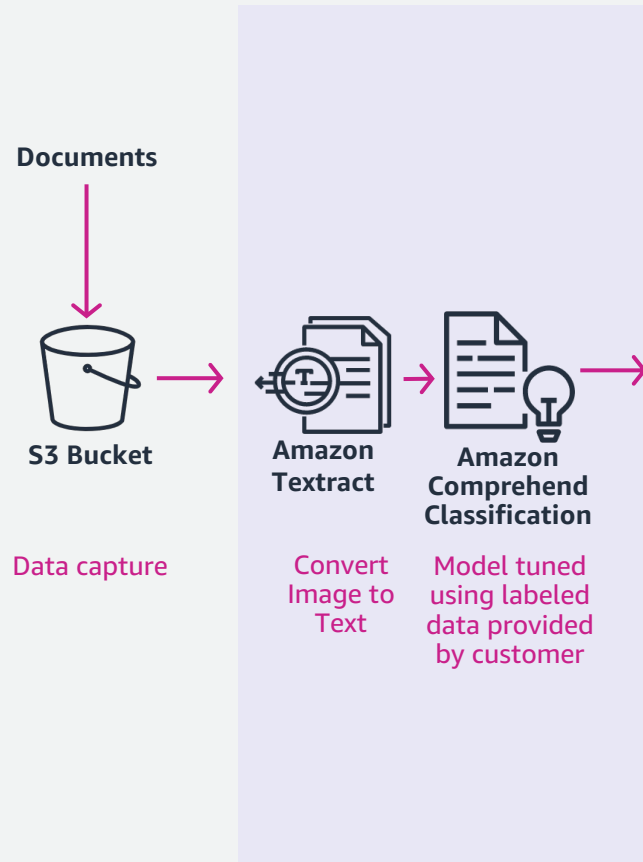
4

Important Messages

ABCDEFGHIJKLMNPNOPQRSTUVWXYZ7890123456789012345678  
ABCDEFGHIJKLMNPNOPQRSTUVWXYZ7890123456789012345678  
ABCDEFGHIJKLMNPNOPQRSTUVWXYZ7890123456789012345678  
ABCDEFGHIJKLMNPNOPQRSTUVWXYZ7890123456789012345678  
ABCDEFGHIJKLMNPNOPQRSTUVWXYZ7890123456789012345678  
ABCDEFGHIJKLMNPNOPQRSTUVWXYZ7890123456789012345678  
ABCDEFGHIJKLMNPNOPQRSTUVWXYZ7890123456789012345678  
ABCDEFGHIJKLMNPNOPQRSTUVWXYZ7890123456789012345678  
ABCDEFGHIJKLMNPNOPQRSTUVWXYZ7890123456789012345678  
ABCDEFGHIJKLMNPNOPQRSTUVWXYZ7890123456789012345678

6

# Intelligent Document Processing pipeline



Categorize, Tag documents  
with Classification

# Aggregating and storing documents



# Document classification

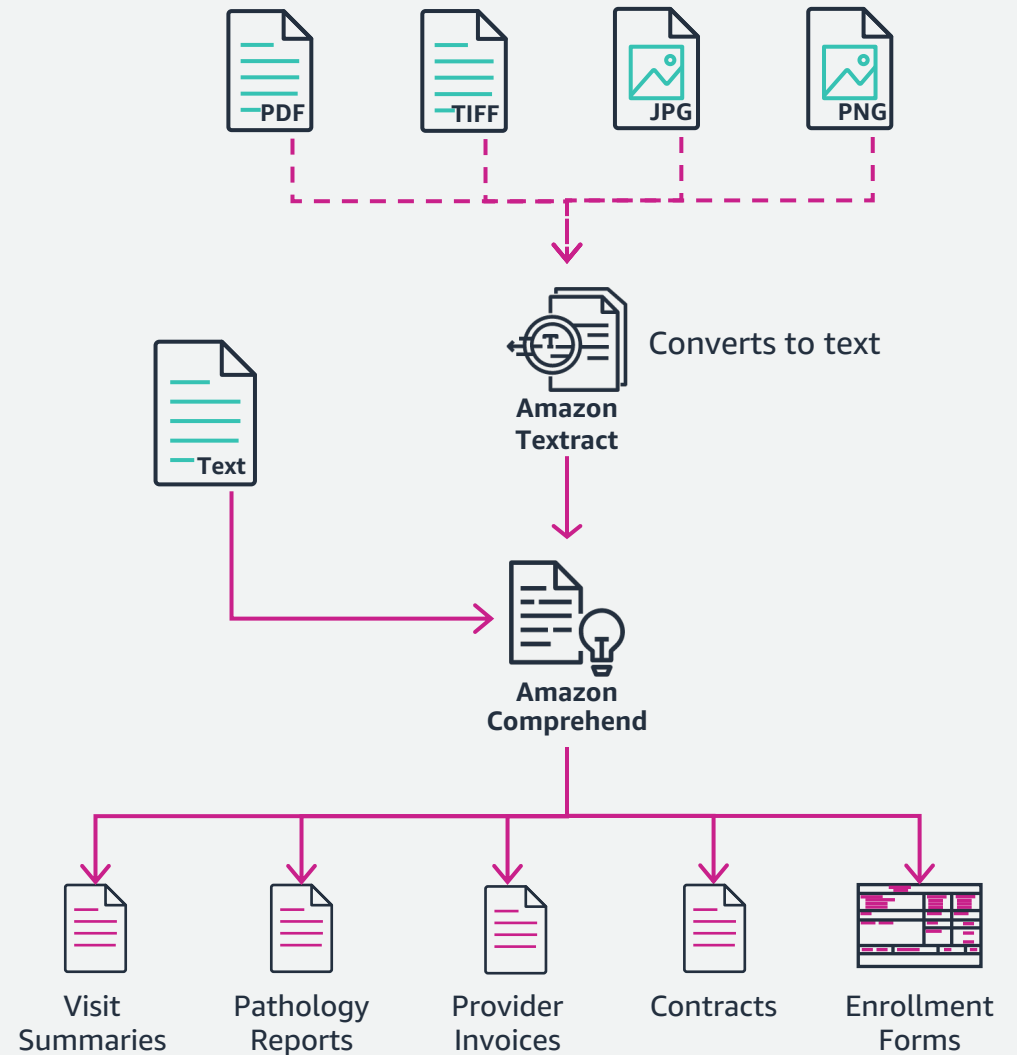
## Identify document types

Sort different types of documents into the right category with Amazon Comprehend **Custom Classifier**, and send them to the correct document pipeline.

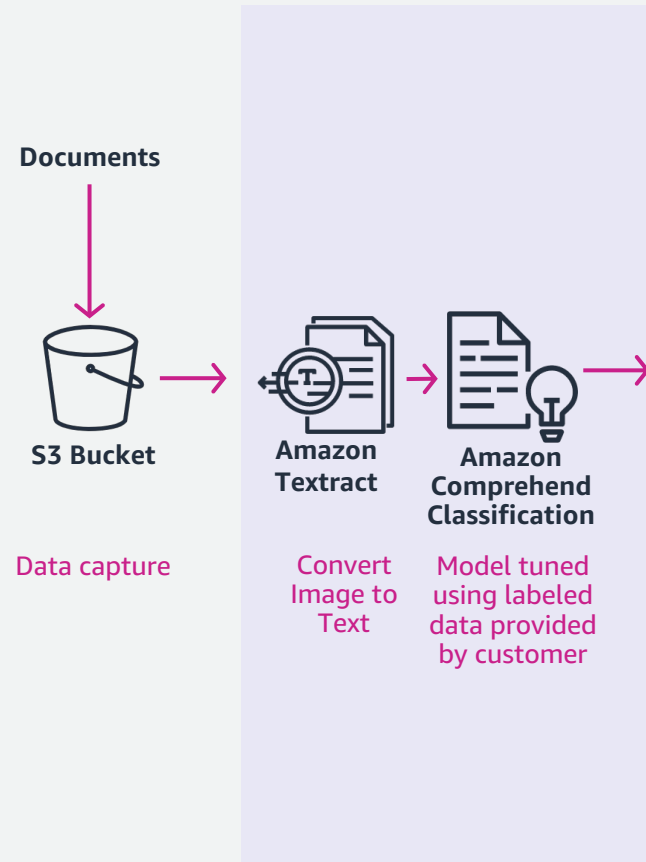
To train a **custom classifier** you need to provide 10-50 samples of each type of document to teach the IDP AI how to classify your documents.

## Real-time or asynchronous processing

Perform real time document classification when needed or reduce costs using asynchronous processes.



# Intelligent Document Processing pipeline



Categorize, Tag documents  
with Classification



# IDP enrichment with Amazon Comprehend

## Redaction of sensitive data

### DOCUMENT TEXT

Hi, my name is John Doe. For verification, the last 4 digits of my social are 6789 and my DOB is 01/01.

Hi, my name is [NAME]. For verification, the last 4 digits of my social are [SSN] and my DOB is [DATE\_TIME].

## Recognize common entities

### DOCUMENT TEXT

Budget Mutual insurance Company  
9876 Infinity Ave  
Springfield, MI 65541  
Client Laura McDaniel and date of death is 6/15/20202

### OUTPUT

**Organization:** Budget Mutual insurance Company

**Location:** 9876 Infinity Ave  
Springfield, MI 65541

**Date:** 6/15/2022

## Recognize entities specific to your business

### DOCUMENT TEXT

Hi my name is John. I am filing a claim for a car accident, my insurance number is 456-YQT.

### OUTPUT

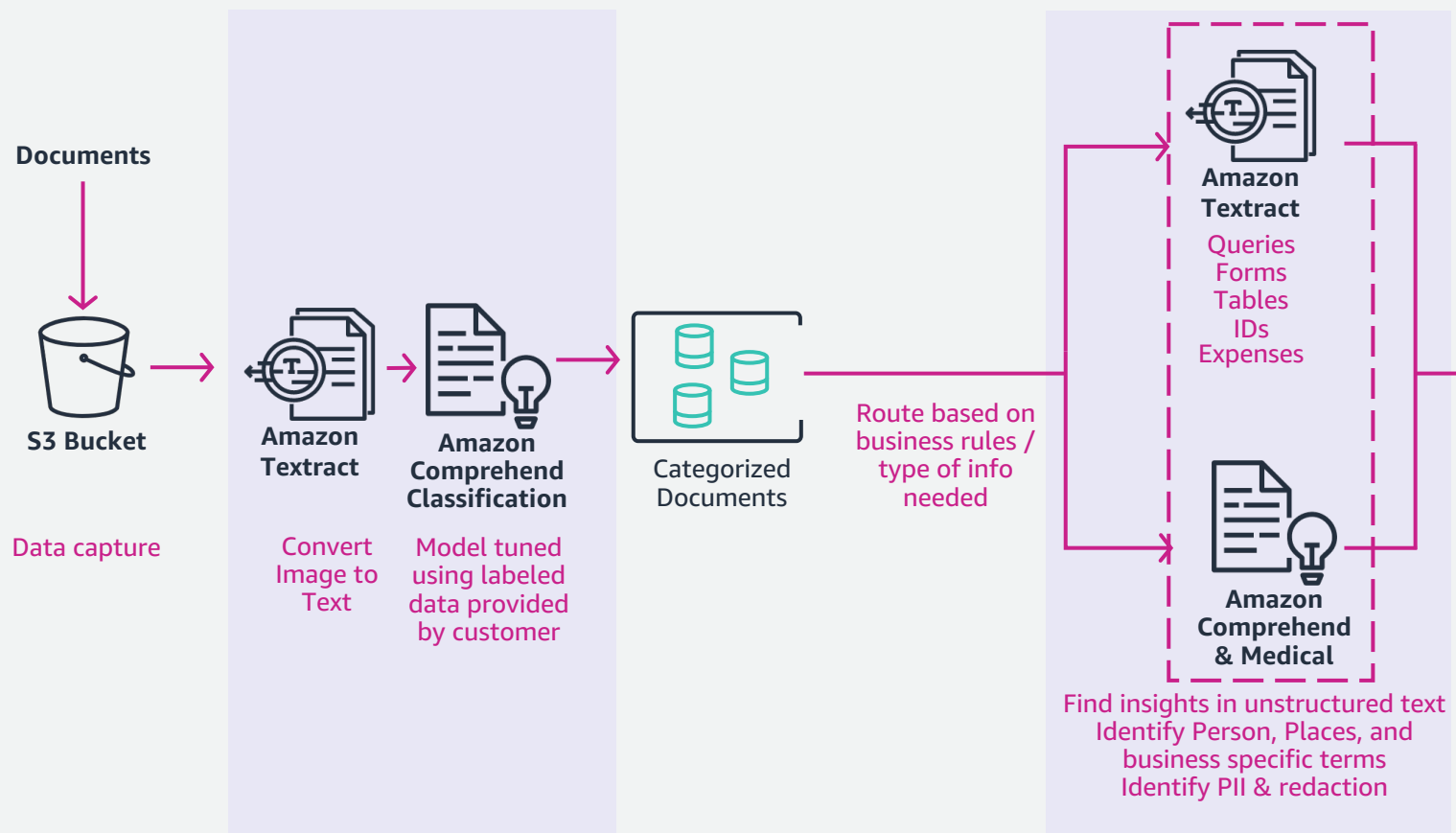
Summary fields:

**Category of entity:**  
Policy\_ID

**Entity:** 456-YQT

**Confidence:** 0.95

# Intelligent Document Processing pipeline



Categorize, Tag documents  
with Classification

Get key insights with  
Extraction & Enrich

# Use a human in the loop when needed



Human  
Review

## Business rules drive when to review

- Example: review if low field accuracy or if claim value is >[\$25K]
- Flexible to use your staff or your designated 3<sup>rd</sup> party
- Can use multiple reviewers for most critical data points
- Admin panels allow business users to update rules

Review the key-value pairs to the right and correct them if they do not match the document below.

Employment Application

Application Information

Full Name: Jane Doe

Phone number: 550-0100

Home address: 123 Any Street, Any Town, USA

Mail address: same as home address

Sample

### Key-value pairs to review

Key-value pair ☒ Yes ☐ No

Full name: ☐ Key is not found

Jane Done

☐ Value is blank

Key-value pair ☒ Yes ☐ No

Phone number: ☐ Key is not found

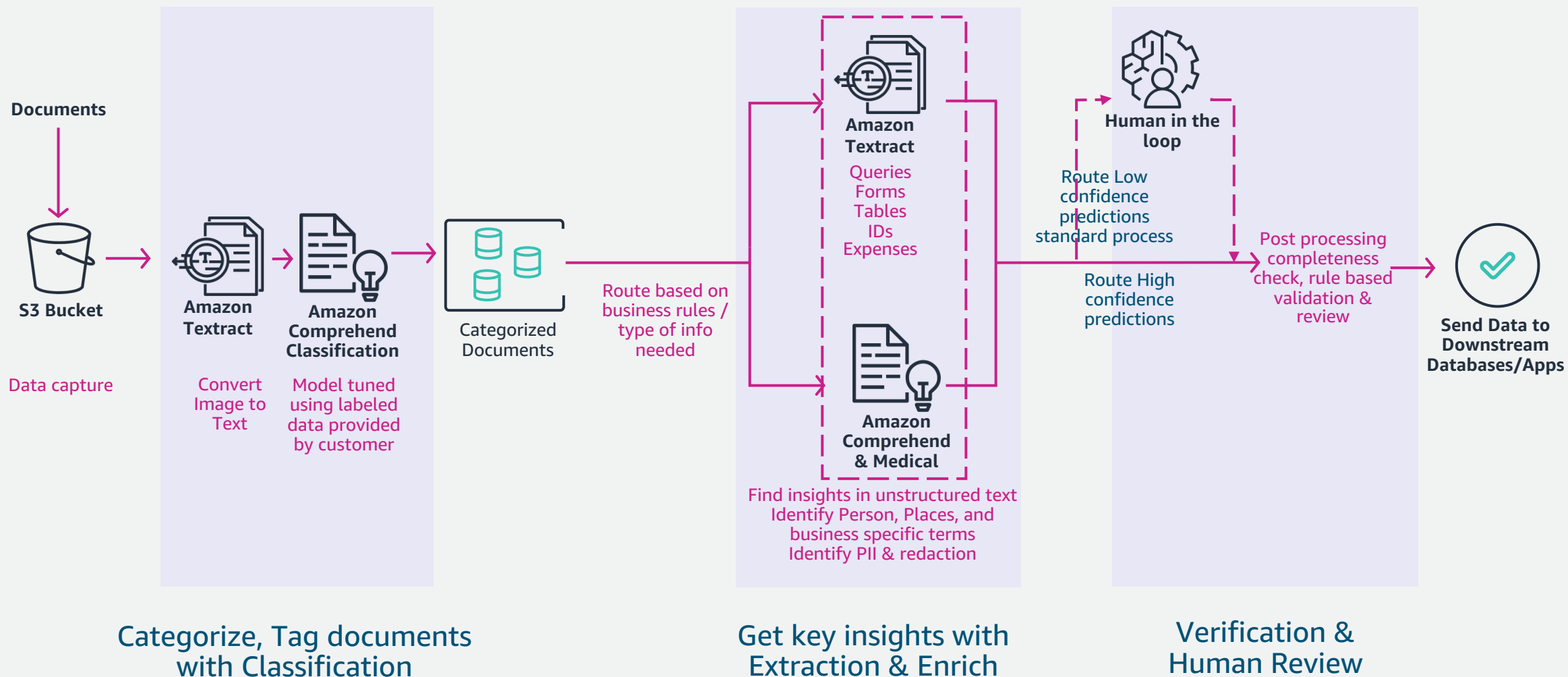
550-0100

☐ Value is blank

☐ No adjustment needed

Submit

# Intelligent Document Processing pipeline



# IDP security, governance, and compliance posture

## Security and governance



### Data privacy and access management

Manage user, access permissions, and policies with IAM  
Encrypt data end-to-end in transit, and at rest with AWS KMS



### Network security and isolation

Isolate data to your private network and region using a Virtual Private Cloud and establish private connections with VPC Endpoints



### Governance, audit, and log management

Enable governance, compliance, and operational and risk auditing with Amazon CloudTrail. Centralized logging with Amazon CloudWatch



## Compliance



ISO 9001/27001/27017/27018



Health Insurance Portability and Accountability Act



Federal Risk and Authorization Management Program



Service Organization Certificate



Payment Card Industry Data Security Standard

# Demo

## Insurance Claims processing



# Improving loan underwriting process productivity



## Problem

Incomplete loan packages, missing data, and the discovery of additional documentation requirements during the underwriting process often creates more work and slows the process down, resulting in fewer loans reviewed by underwriters and added lender costs.



## Solution

Black Knight developed Underwriter Assist (UA), an AI-powered mortgage solution that leverages Amazon Textract and other machine learning technologies. UA performs document recognition, data extraction, and analysis of income, assets, and property appraisal.



## Impact

Using advanced machine learning services like Amazon Textract, Black Knight has increased lender productivity, driven operational efficiency and reduced costs, resulting in higher customer satisfaction.





# Scaling claims processing through automation



## Problem

Health insurance companies spend millions of dollars to extract sensitive information from claims forms and accompanying attachments to perform their business operations. To reduce manual labor, the company wanted to automate the process.



## Solution

Elevance Health chose Amazon Textract to digitize and automate its claims process, for its image-processing capability, ability to detect tables and forms, and adherence to security and compliance standards.



## Impact

Using Amazon Textract, Elevance Health can extract and digitize data to quickly process thousands of claims each day. Elevance Health has automated classification of attachments for claims process by 90%





# Skills required to implement Intelligent Document Processing

## Connects IDP to existing workflow

- Serverless development
- Moves documents through IDP workflow and sends information to downstream systems
- Learns how to use AI APIs

AWS Developer

Project Manager

Business Systems Analyst

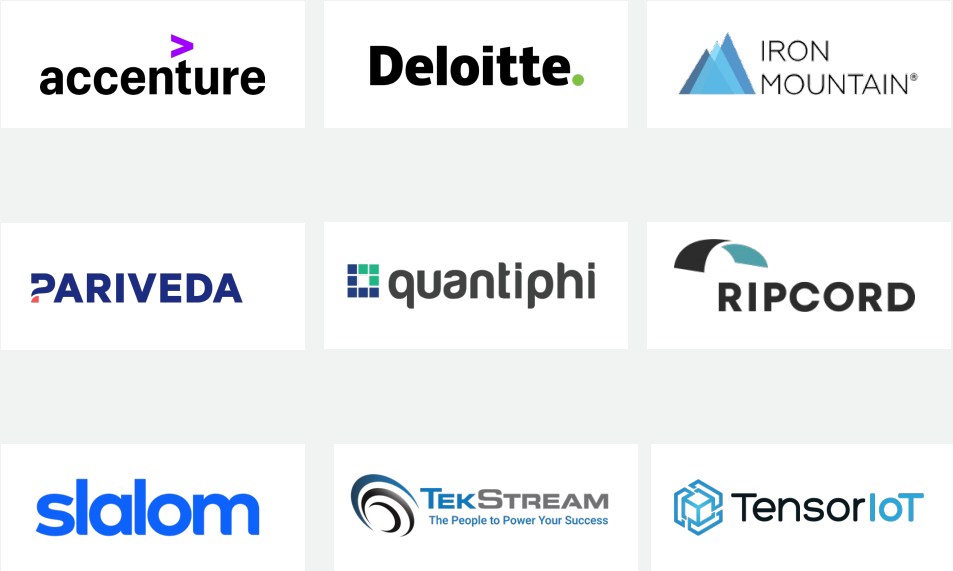
## Stakeholder coordination

- Drives towards executive sponsor's goals
- Ensures upstream / downstream workflow stakeholders are aligned

## Ensures document workflow meets business needs

- Owns IDP roadmap
- Sets accuracy targets
- Provides labeled examples for document classification and custom business entities

# Amazon partners can help accelerate IDP projects



	Suggest Build In-House	Suggest Partner
Document processing is a top corporate competency	X	
Difficulty hiring enough experienced AWS developers		X
Using a partner for digital transformation projects		X
Already use serverless architecture to manage document workflows	X	

**Up next...**

**IDP Lab**

