
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

AIM334

Using AI to process healthcare data with intelligent document processing

Sonali Sahu

Sr. AI Services Specialist SA Leader, IDP
AWS

Anjan Biswas

Sr. Specialist Solutions Architect, IDP
AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

- Digital transformation with AI/ML
- Use case
- Intelligent document processing (IDP) workflow
- AWS AI services used in IDP
  - Amazon Textract
  - Amazon Comprehend
  - Amazon Comprehend Medical
- Solution design
- Demo
- Resources

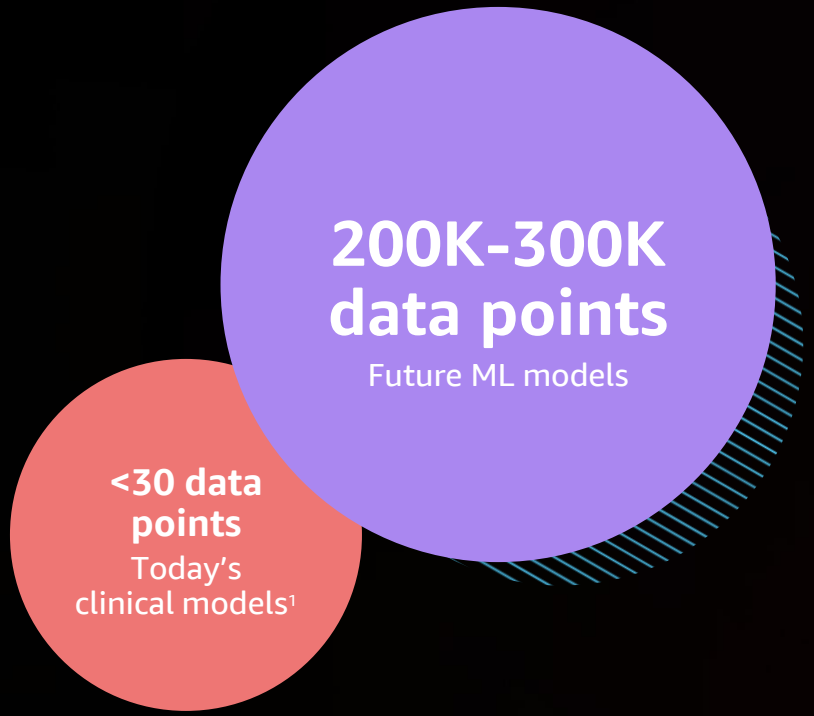
ASK QUESTIONS AND SPEAK OUT!

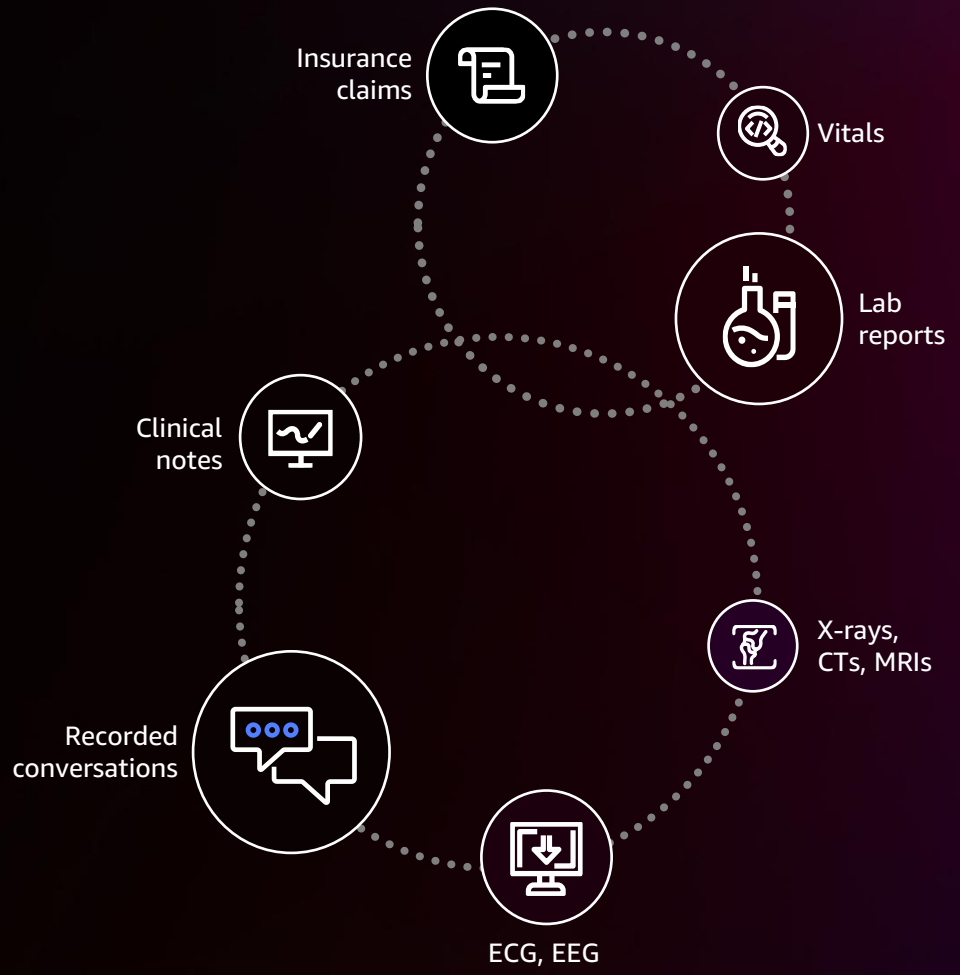
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Digital transformation

HEALTHCARE AND LIFE SCIENCES





Health data is frequently unstructured, multimodal, and siloed

Information extraction is labor-intensive and error-prone

Difficult to derive actionable and meaningful insights

1. Improving Hospital Mortality Prediction with Medical Named Entities and Multimodal Learning, https://arxiv.org/abs/1811.12276

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Use case overview

ANYCOMPANY HEALTHCARE CORP. (AHC) – AUTOMATING HEALTH INSURANCE CLAIMS PROCESSING

- **Health insurance paper claim document packages**
  - >3,000,000 claims per year
  - Peak 15,000 claims a day
- **Need to ensure PII and PHI protection**
- **Need de-identification of documents per HIPAA privacy rule**
- **Need to process as many claims as possible – automated, without human intervention**



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Intelligent document processing (IDP)

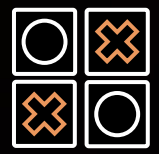
HIGH-LEVEL END-TO-END WORKFLOW – HEALTH INSURANCE CLAIMS PROCESSING

Data capture



Centralized location to store all documents

Classification



Identify and label documents

Extraction



Extract key information and entities from documents

Enrichment



Enrich documents – Redact, de-identify, tag documents

Review and validation



Route data with poor accuracy to a human to review; run business rules

Ready for business decisions



Update claims database with all information; run analytics on the data

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

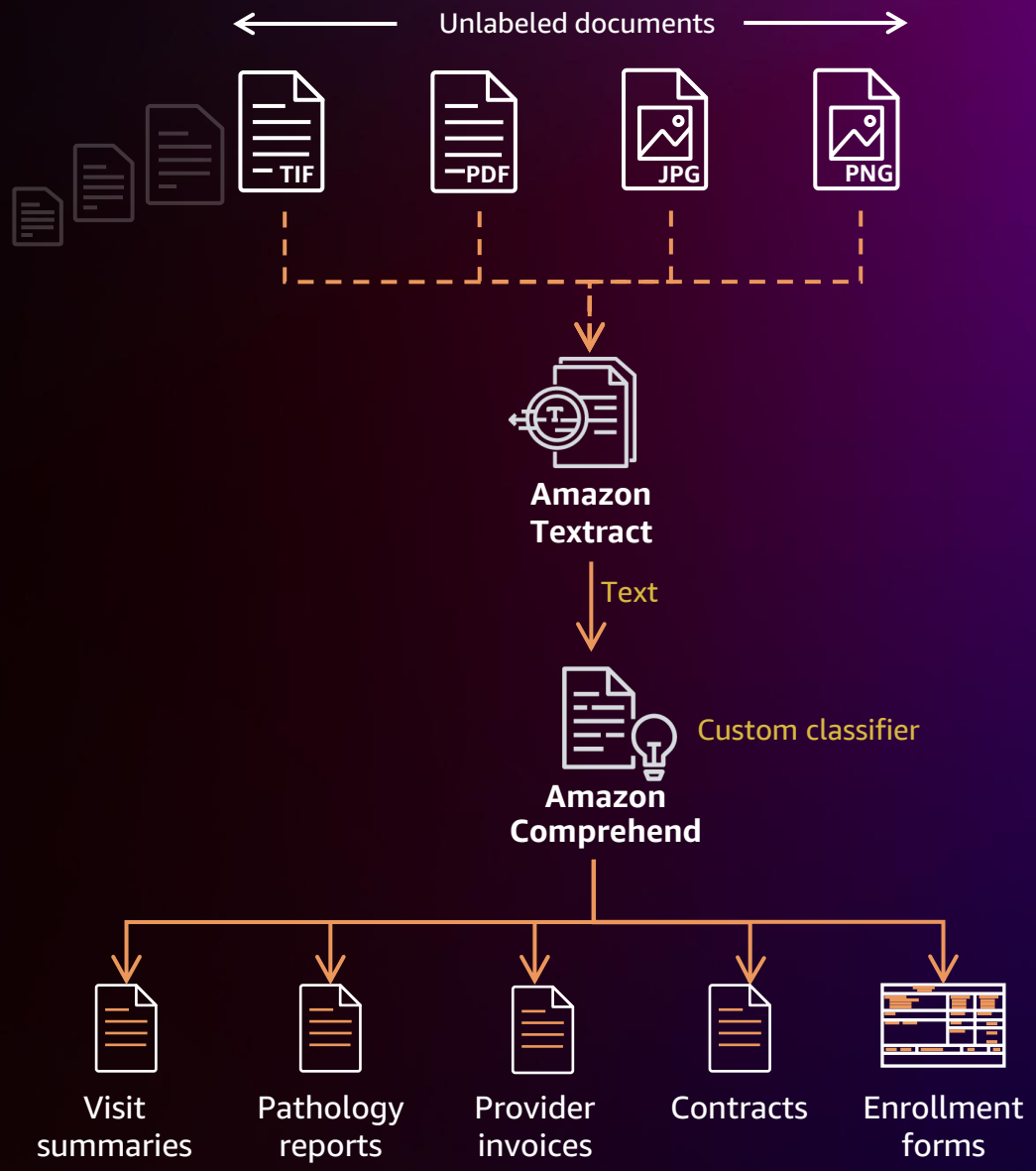
Document classification

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

Real-time or asynchronous processing

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



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data extraction using Amazon Textract

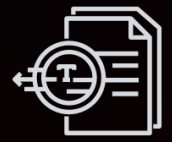
FROM STRUCTURED, SEMI-STRUCTURED, AND UNSTRUCTURED DOCUMENTS



Documents

Physician notes, medical charts, lab reports, transcripts, and more

PDF, PNG, JPEG, & TIFF file formats



Amazon Textract

Text
Lines and words

Forms
Key-value pairs

Tables
Headers, rows, cells, and indexes

Handwriting
Handwritten text

Queries
Search specific info from documents

Invoice and receipt
Normalized summary and line items

ID documents
Normalized data from US identity documents – passports, driver's licenses, ID cards

Geometry of items detected
Bounding box and polygon

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data extraction using Amazon Comprehend



Documents

Physician notes, medical charts, lab reports, transcripts, and more

PDF/PNG/JPEG

Amazon Comprehend

Extract insights from text

Amazon Textract

Text

Identify common entities and custom entities

Common entities

- DATE
- EVENT
- LOCATION
- ORGANIZATION
- PERSON
- QUANTITY
- TITLE

Custom entities

- PATIENT_ID
- POLICY_ID
- ADMIT_DATE
- PROVIDER_ADDRESS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Extract medical insights and detect PII/PHI

WITH AMAZON COMPREHEND MEDICAL AND AMAZON COMPREHEND



Documents

Physician notes, medical charts, lab reports, transcripts, and more

Amazon Textract → Text → **Amazon Comprehend Medical**

Automatically extract medical /clinical insights from text

Identify relationships among extracted health information and link to medical ontologies

- SNOMED CT linking
- RxNorm
- ICD-10-CM linking
- PHI entities

Amazon Textract → Text → **Amazon Comprehend**

Detect PII

Identify and extract personally identifiable information (PII)

Name	Email
Address	Age
Phone	SSN

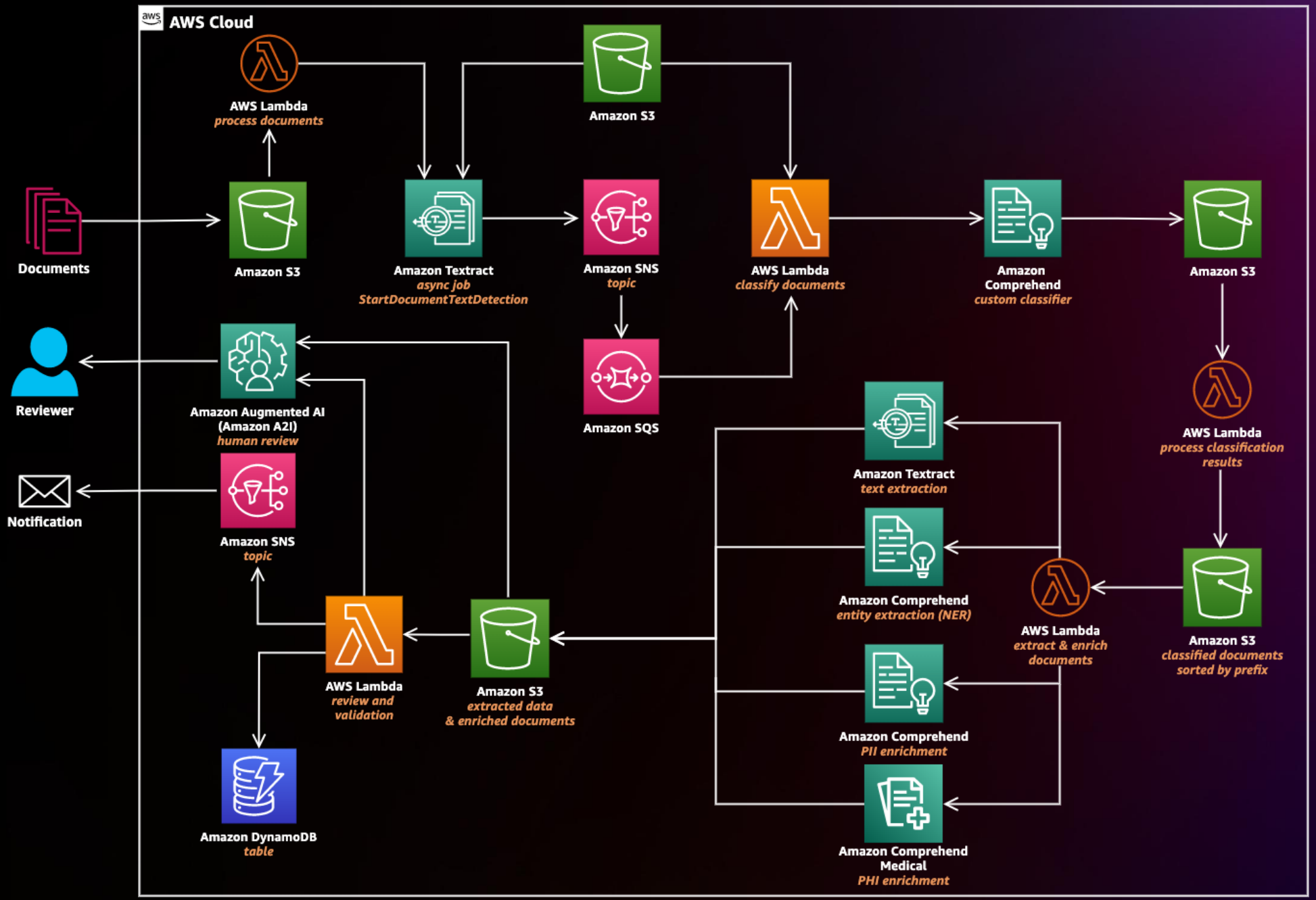
And more

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Intelligent document processing

HEALTH INSURANCE CLAIMS PROCESSING SOLUTION DESIGN

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Resources

Getting started resources

Open-source workshop

IDP CDK Constructs

Open-source de-identification app

Scan the QR codes to access resources and open-source code

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Sonali Sahu
sonalsah@amazon.com

Anjan Biswas
anjanavb@amazon.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.