

USE201

# Streamline document processing with AI/ML

Jihun Kim

Senior Data/ML Architect,  
Professional Services

AWS

Chris Suter

Lead Infrastructure and  
Connectivity Architect

NHSX



# Agenda

- Overview of document processing
- Amazon Textract and Amazon A2I
- Streamlined text extraction with Amazon Textract, SageMaker, and Amazon A2I
- Customer story: Automated document processing at NHSBSA

# Document processing

UPWARD OF 80% OF ENTERPRISE DATA TODAY IS UNSTRUCTURED\*

Documents are still primary business tools across industries

Lack of effective tools and technologies to extract business values



\*Source: [Gartner](#)

# How documents are processed today



**Manual processing**

**Expensive**

**Error prone**

**Time consuming**

---



**Optical Character Recognition (OCR)**

**Simple documents only**

**Error prone**

**Dump of text**

---



**Rules and template-based extraction**

**Limited by accuracy of OCR**

**Development and management overhead**

**Templates are brittle**

# Amazon Textract + Amazon A2I

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



**No ML experience  
required**

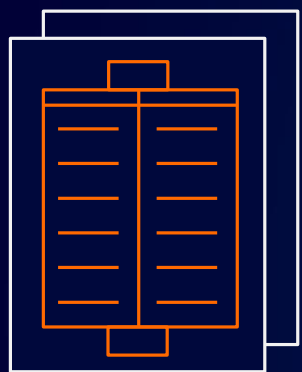
## Benefits of **Amazon Textract**

**Eliminate  
manual effort**  
(including template  
development)



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

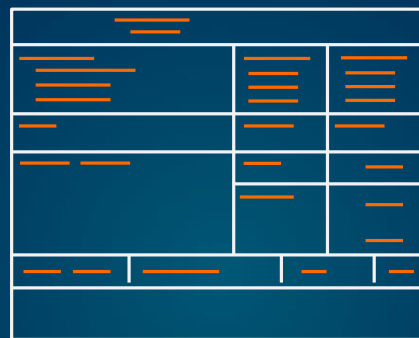
# Amazon Textract features



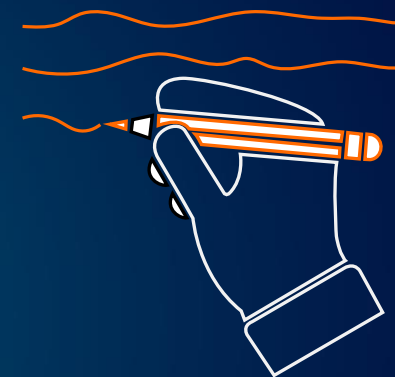
Text  
Extraction



Table  
Extraction



Form  
Extraction



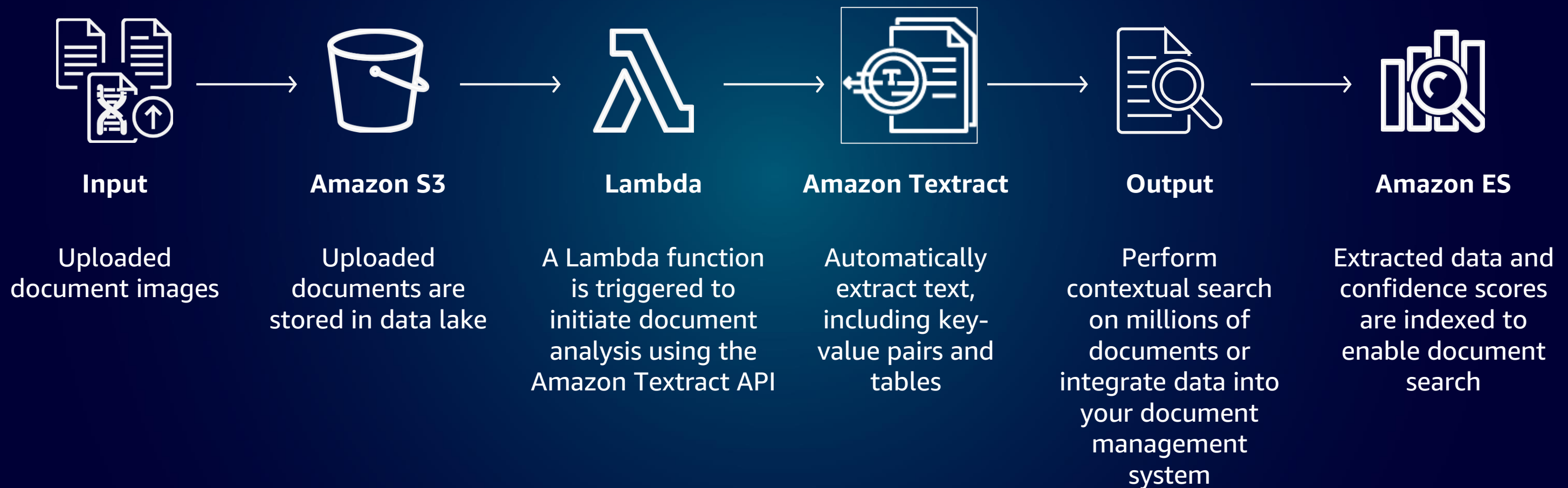
Handwriting



Human in the Loop  
(Amazon A2I)

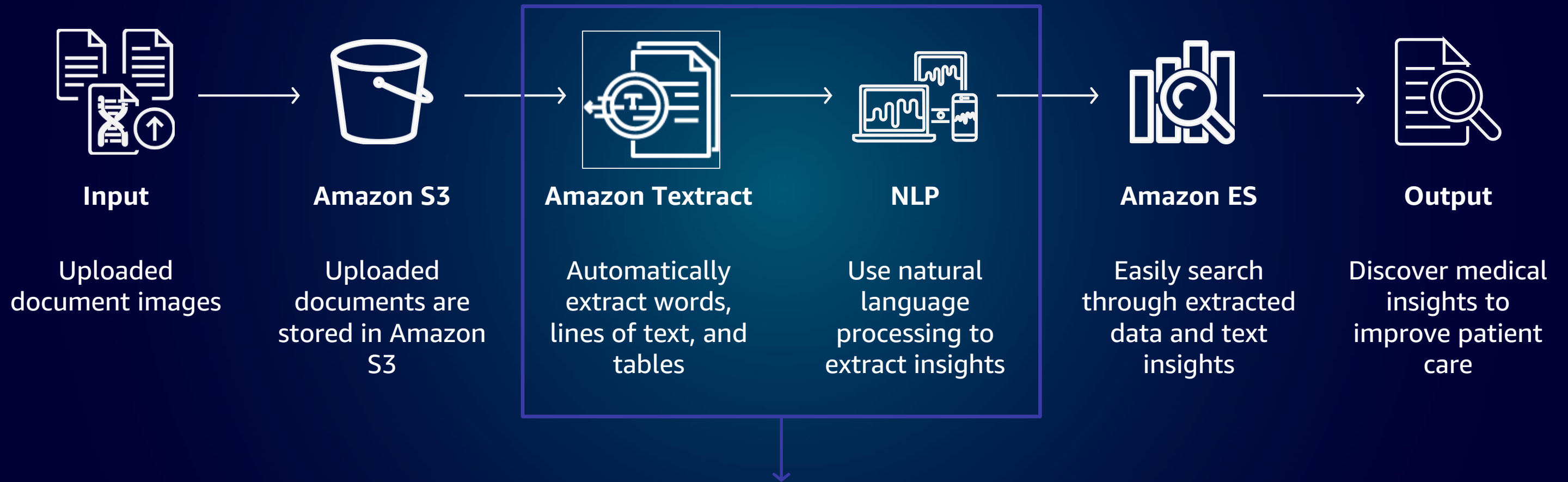
# Reference architecture

## INDEX AND SEARCH DOCUMENTS



# Reference architecture

## EXTRACT FOR NLP



Quickly turn extracted text/data into actionable insights

# Textract works with Amazon A2I for human review

Regulatory requirements

Sensitive decisions

Hard-to-read documents

...more



Human review provides oversight and review for low-confidence results

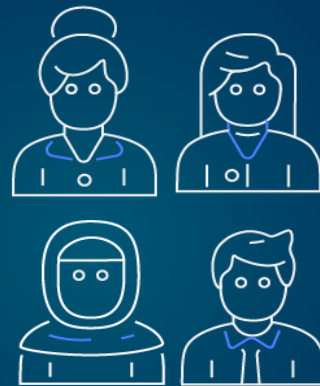
# Benefits of Amazon A2I



Easily implement human review workflows



Reduce time to market with pre-built workflows and UIs



Multiple workforce options



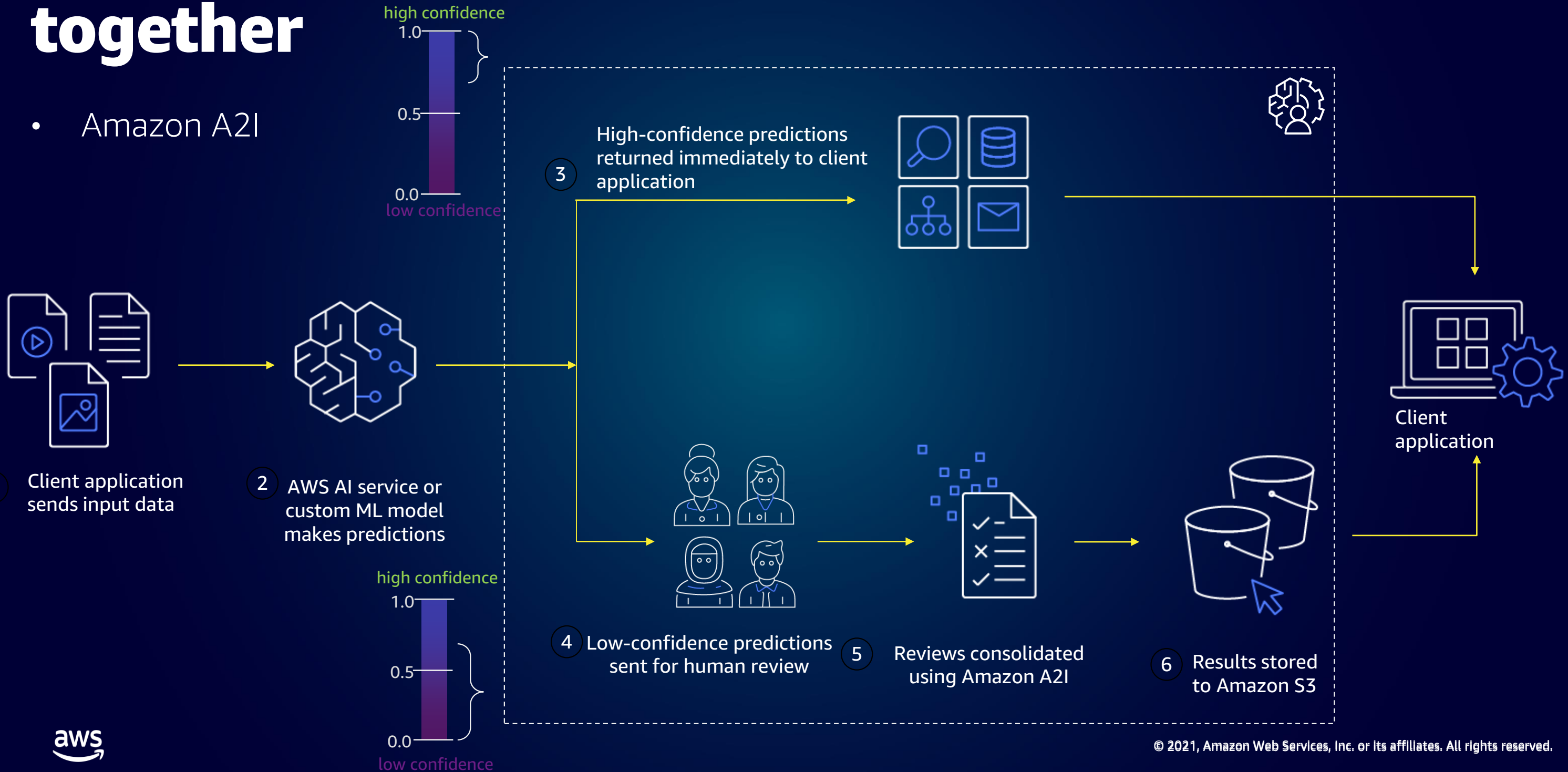
Integrate with your custom ML models



Increased worker efficiency

# How Amazon Textract and Amazon A2I work together

- Amazon A2I



# Demo summary

## AMAZON Textract

### Amazon Textract and Amazon A2I

- Documents uploaded to a Amazon S3 bucket
- Extracted data from the sample form
- Handwriting recognition

### Amazon A2I

- Incorporate human reviewing process to ML prediction
- Review the output (before & after)



**Business Services Authority**



## National Health Service

- UK government funded medical and healthcare service
- Everyone in the UK can use
- Free at the point of use





**66M** Population

---

**1M** Patients every 36 hours

---

**7K** Medical practices

---

**350+** Hospital trusts and care groups

# SUPPORTING THE NHS

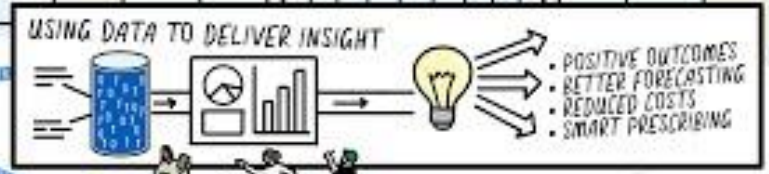


NATIONAL SCALE, LOCAL DELIVERY



AUTOMATED CHECKING

CASHFLOW



£35bn TRANSACTIONS

TOTAL AMOUNT OF MONEY FLOWING THROUGH OUR ORGANISATION

OUR LEVERS

OUR PEOPLE

CITIZEN SERVICES

DIGITISATION

STRATEGIC SOURCING

BUSINESS PROCESS IMPROVEMENT

EMERGING TECHNOLOGIES

DATA

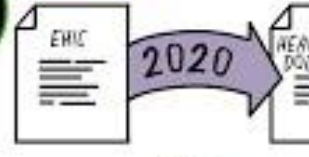
OUR LEVERS

RELATIONSHIPS AND NETWORKS

WORKFORCE SERVICES

PRIMARY CARE SERVICES

5.5 MILLION  
EMICS ISSUED ANNUALLY



PUBLIC FACING SERVICES

£400m





## Business Services Authority

**£35B** Budget

---

**£722M** Savings since 2013

---

**1.4M** NHS employees managed

---

**1B +** Prescriptions processed



# Business Services Authority

**45M**

Prescriptions per month

**£9.4B**

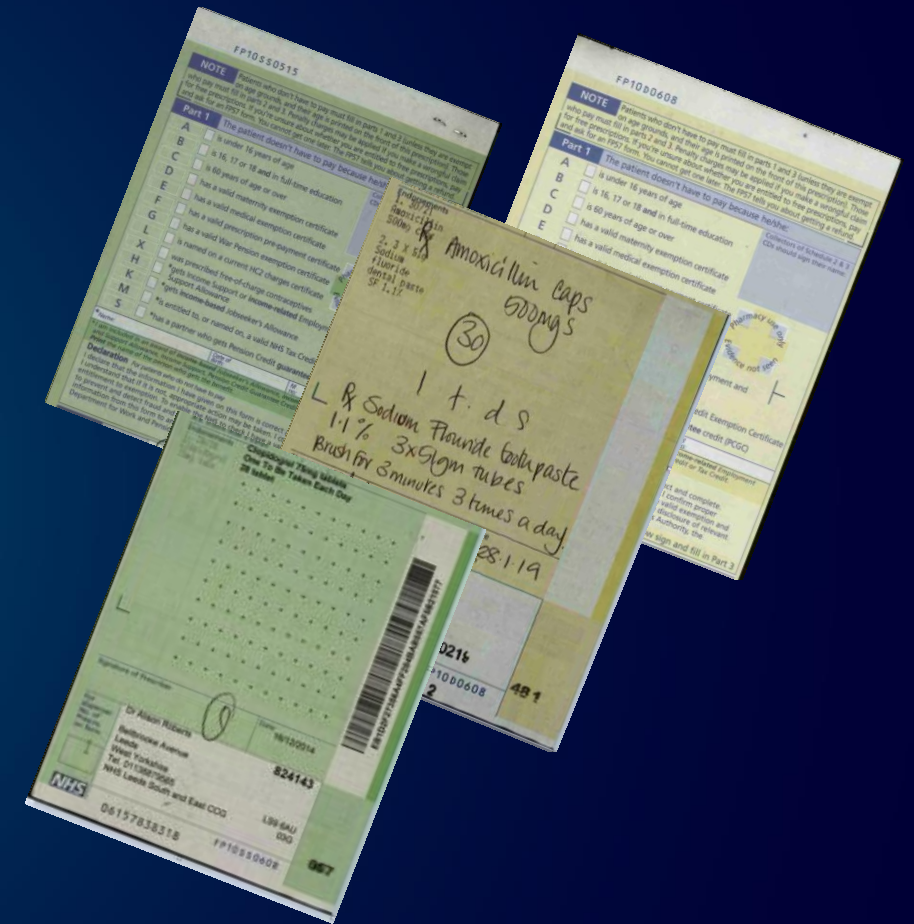
Reimbursed

**18M**

Paper based prescriptions

**2.4**

Prescribed items per form



# Challenges...

- Overall scale
- High operational costs
- Diversity and variation of prescriptions
- Handwriting
- Ease of maintenance
- Other use cases

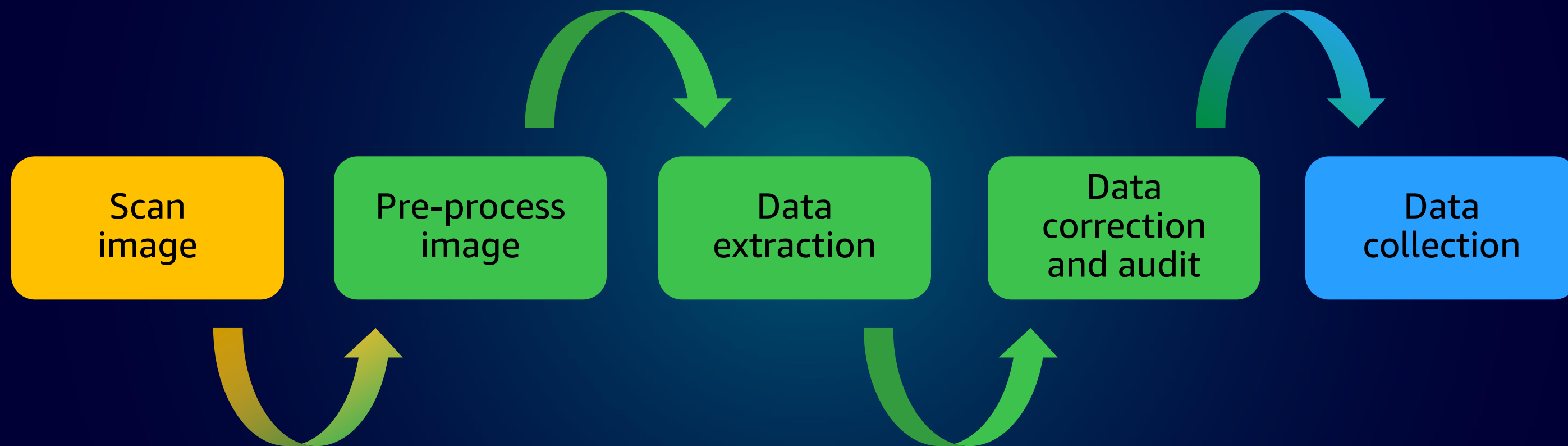


# Requirements...

- Increase read rate accuracy through automation
- Reduce human effort and cost per transaction
- Extract more data to improve process and insights
- Develop a self managed capability to allow the same approach on other paper based activities within the NHSBSA
- Off the shelf – with little specialist maintenance as possible



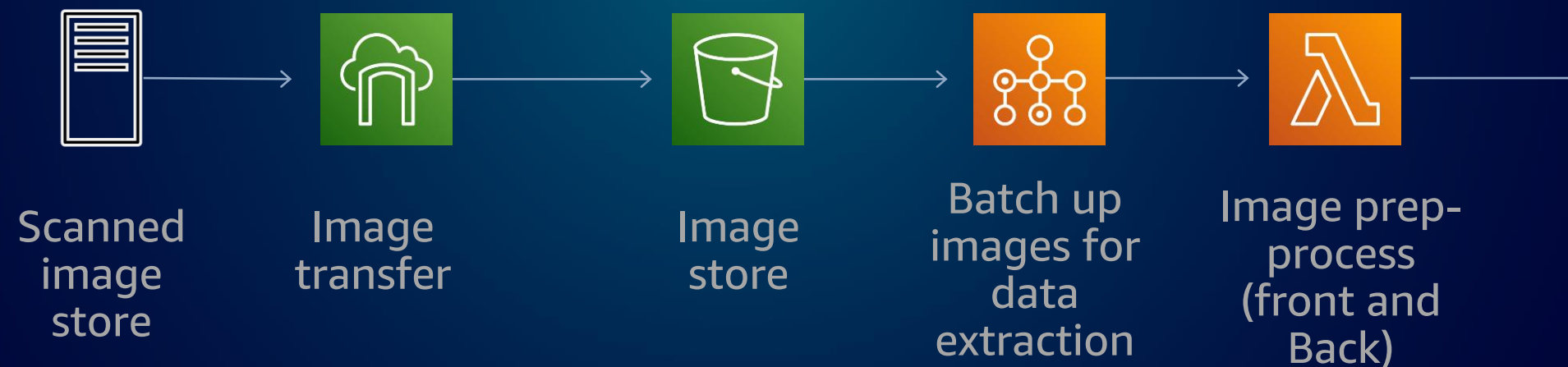
# Processing flow...



# Pre-process image...

## Part 1

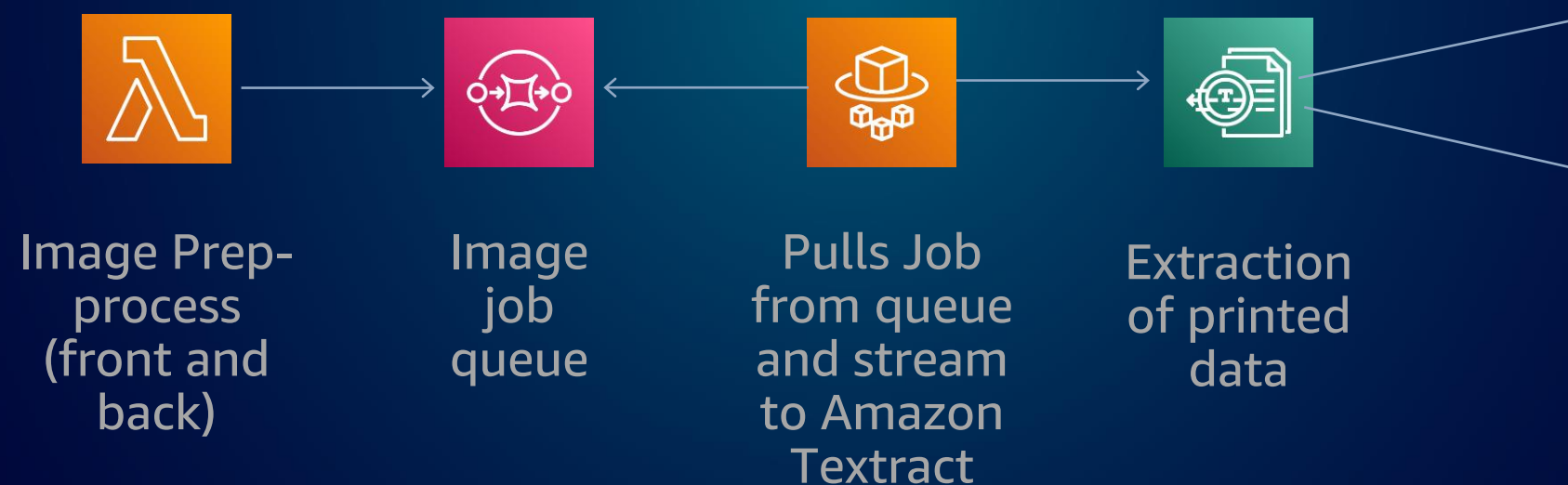
- Getting images from on-premises to AWS
- Group scanned images (front and back)



# Extraction of printed data...

## Part 2

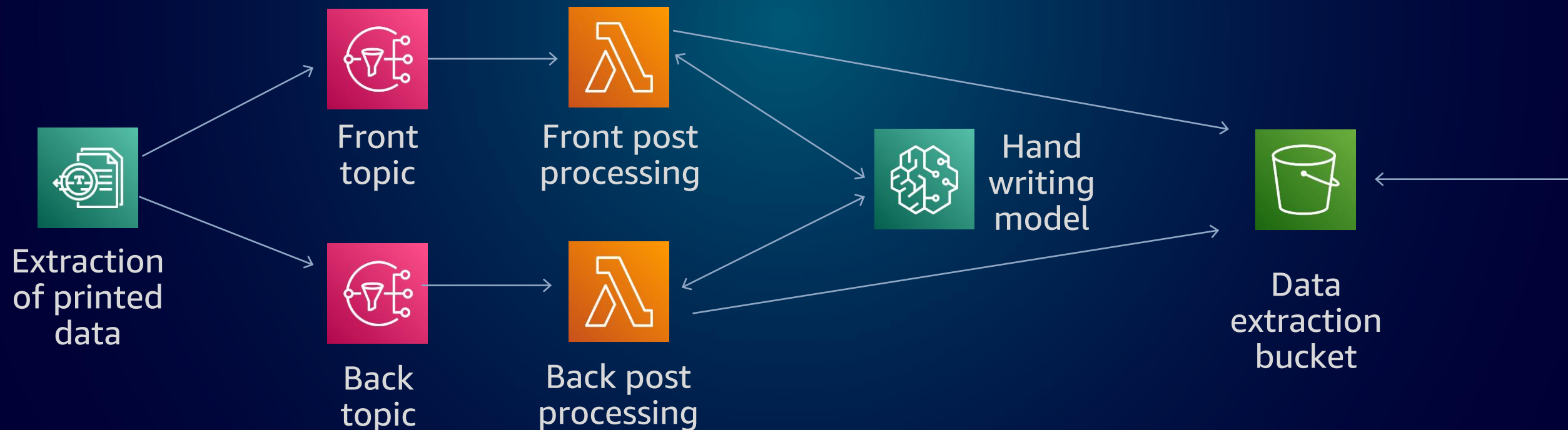
- Extraction of printed data from image
- Amazon Textract API call async



# Extraction of hand-written data...

## Part 3

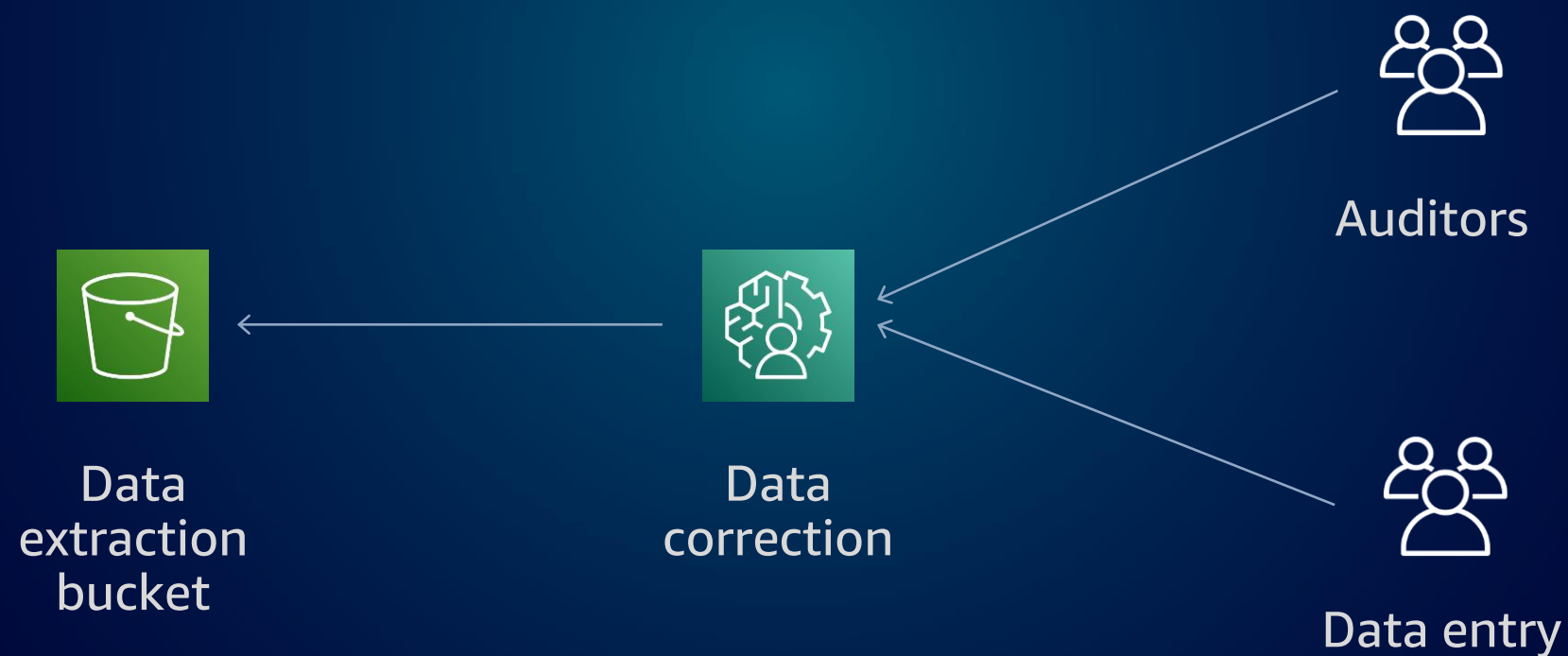
- Recognition of handwritten data within images
- SageMaker handwritten model (before Amazon Textract Handwritten)



# Data correction and audit...

## Part 4

- Understanding image confidence scoring
- Human data correction using Amazon A2I



# Solution summary...

## Complete serverless architecture



# Achievements..

- Overall scale
- Low operational costs
- Flexibility
- Diversity and variation of prescriptions
- Handwriting
- Ease of maintenance



**“Success is not final. Failure is not fatal. It is the courage to continue that counts.”**

**– Winston Churchill**

# Possible future landscape...

- Indexing of healthcare information and data
- Fraud and error detection
- Insights
- More form types
- Continued ML retraining from Amazon A2I outputs

# Learn ML with AWS Training and Certification

LEARN TO APPLY MACHINE LEARNING TO YOUR BUSINESS, UNLOCKING NEW INSIGHTS AND VALUE



Courses appropriate for your role, including developers, data scientists, data engineers, and business decision makers



65+ no-cost digital machine learning courses from AWS experts let you learn from real-world challenges solved by Amazon



Build credibility and confidence with AWS Certification, including *AWS Certified Machine Learning – Specialty*



Go deeper with labs, whitepapers, tech talks, and more by accessing the [AWS Ramp-Up Guide](#)

Visit [aws.training/MachineLearning](https://aws.training/MachineLearning)

# Thank you!

Jihun Kim

Senior Data/ML Architect  
Professional Services  
AWS

Chris Suter

Lead Infrastructure and  
Connectivity Architect  
NHS X

