



# INNOVATE

DATA AND AI/ML EDITION

**22 February 2023**

# Building an intelligent document processing solution

**Abhijit Kalita**

Senior AI/ML Evangelist, Public Sector Partners  
Amazon Web Services

# Agenda

- Common challenges in document processing
- Overview of Intelligent Document Processing (IDP) framework
- Step 1: Extract text using serverless architecture
- Step 2: Analyze using Amazon Comprehend
- Step 3: Verify – human in the loop
- Key takeaways

# Document processing challenges



Time-consuming/  
Expensive

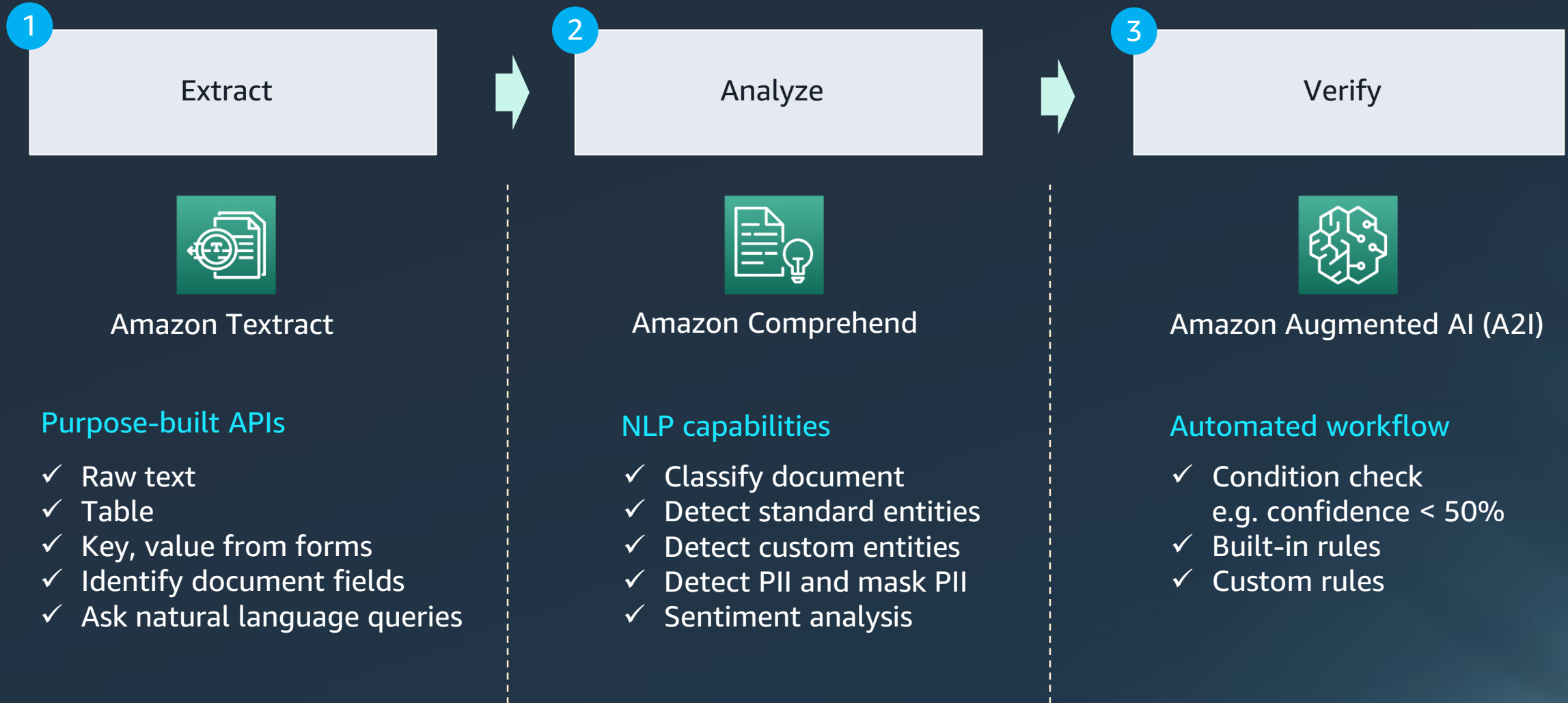


Quality of Optical  
Character Recognition  
(OCR)



Scaling problem

# Intelligent Document Processing (IDP) framework



# Step 1: Extract text using serverless architecture

## Input:

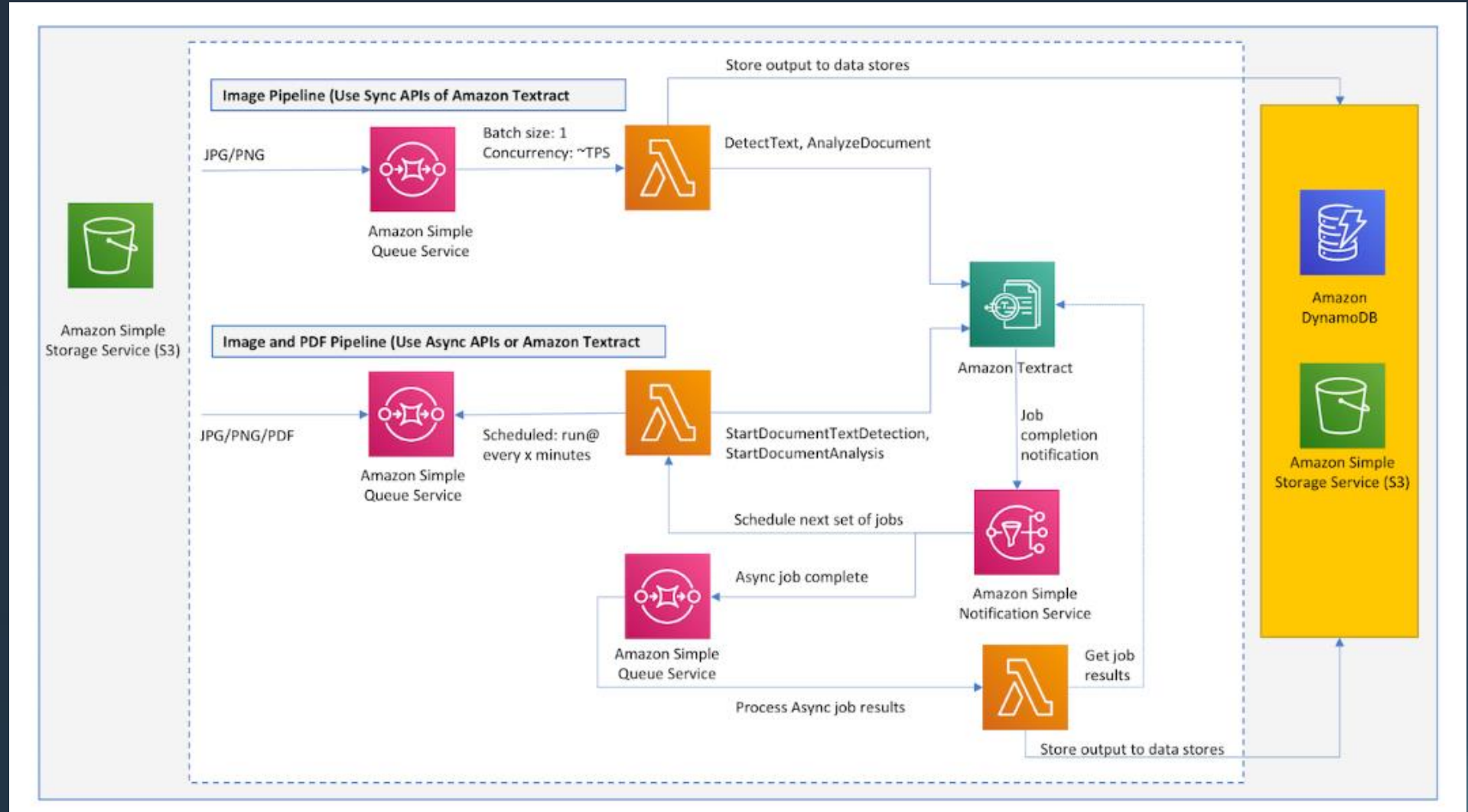
- Drop document in Amazon S3
- PNG, JPEG, TIFF, and PDF

## Output:

- ✓ Raw text
- ✓ Table
- ✓ Key, value from forms (e.g. First Name: Michael)

## APIs:

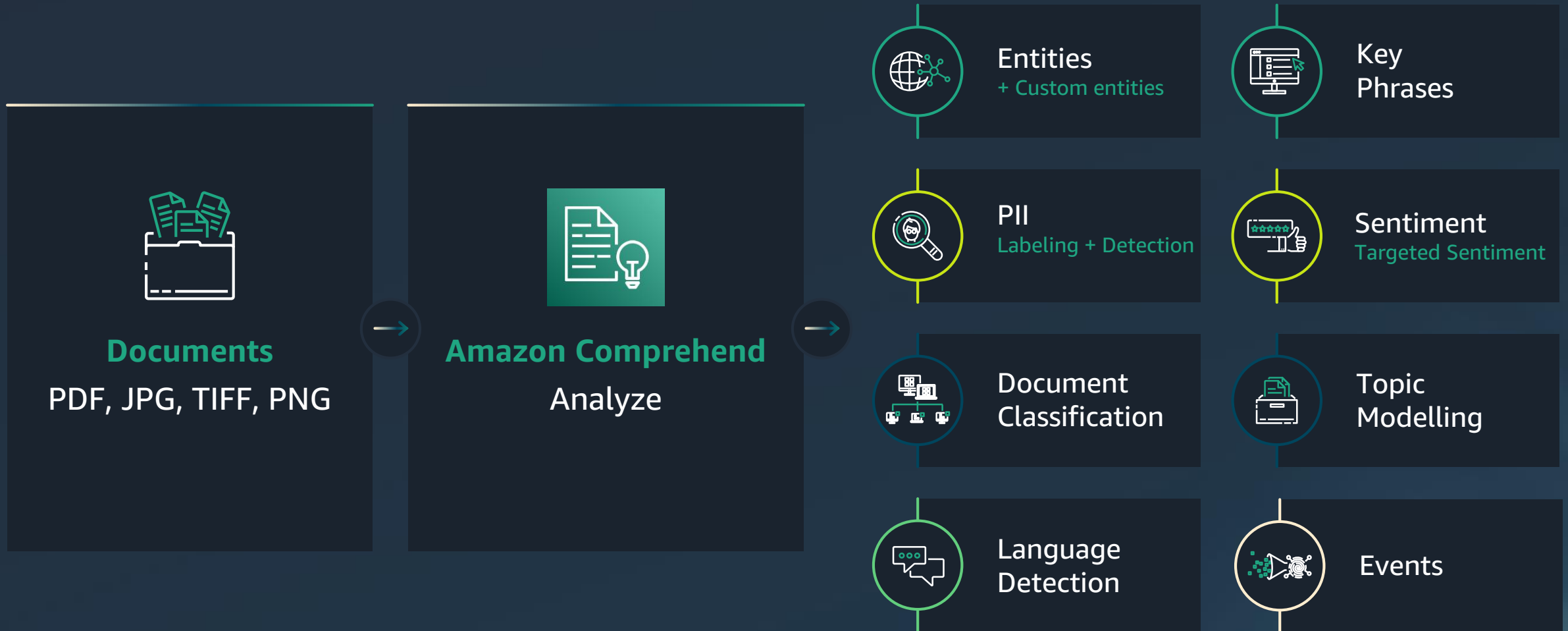
- ✓ Detect document text
- ✓ Table
- ✓ Form



<https://github.com/aws-samples/amazon-textract-serverless-large-scale-document-processing>

# Demo 1: Extract text

# Step 2: Analyze using Amazon Comprehend





# Demo 2:

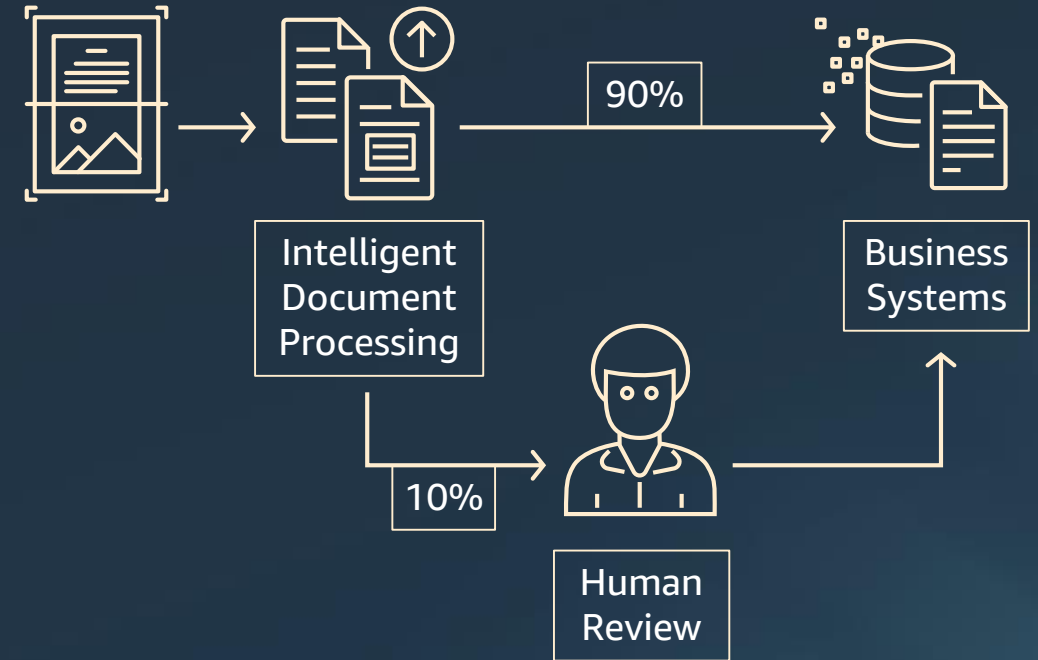
# Document classification

# Step 3: Verify - human in the loop

Amazon Augmented AI (A2I) provides **built-in human review workflows**.

## Criterion to start human review workflow:

- Confidence score < target
  - Key-value pair
  - Text within key-value pair
- Specific key or value missing
- Random sampling
- Custom rules



# Demo 3:

# Web application for IDP

# Key takeaways



Automate text extraction

Operational efficiency



Insights using AI/ML

Create business value



Human in the loop

Continual improvement

# We are here to help you!



**AWS AI/ML Specialists**  
**AWS Professional Services**  
**AWS ML Lab**  
**AWS Data Lab**  
**AWS Prototyping**  
Leverage skills and experience  
from AWS experts



**AWS Partners**  
A global network of software  
and services partners with  
AWS expertise



**AWS Support**  
Engage AWS support team

# Additional resources

Github repo for the serverless stack for OCR

<https://github.com/aws-samples/amazon-textract-serverless-large-scale-document-processing>

2 min video

[Automate data extraction and analytics from documents with machine learning](#)

Blog post

[Building an end-to-end intelligent document processing solution using AWS](#)

Intelligent Document Processing workshop material

<https://catalog.us-east-1.prod.workshops.aws/workshops/c2af04b2-54ab-4b3d-be73-c7dd39074b20/en-US/>

# Visit the Data & AI/ML resource hub

Dive deeper into these resources, get inspired and learn how you can use AI and machine learning to accelerate your business outcomes.

- 6 steps to machine learning success e-book
- 7 leading machine learning use cases e-book
- Machine learning at scale e-book
- Achieving transformative business results with machine learning e-book
- Tackling our world's hardest problems with machine learning e-book
- Accelerating machine learning innovation through security e-book
- ... and more!

**Visit resource hub**



<https://bitly.co/FqdC>



# AWS Training and Certification

Access the AI & ML learning plan courses built by AWS experts on AWS Skill Builder

- Get started with digital self-paced, on-demand training and ramp-up guides to help you grow your technical skills
- Learn how to apply machine learning, artificial intelligence, and deep learning to unlock new insights and value in your role
- Take the steps today, towards validating your expertise with an AWS Certified Machine Learning – Specialty Certification



<https://bit.ly/3FnxDH7>

Learn your way [explore.skillbuilder.aws](https://skillbuilder.aws) »





# Thank you for attending AWS Innovate – Data & AI/ML Edition

We hope you found it interesting! A kind reminder to **complete the survey**.  
Let us know what you thought of today's event and how we can improve the event experience for you in the future.



[aws-apj-marketing@amazon.com](mailto:aws-apj-marketing@amazon.com)



[twitter.com/AWSCloud](https://twitter.com/AWSCloud)



[facebook.com/AmazonWebServices](https://facebook.com/AmazonWebServices)



[youtube.com/user/AmazonWebServices](https://youtube.com/user/AmazonWebServices)



[slideshare.net/AmazonWebServices](https://slideshare.net/AmazonWebServices)



[twitch.tv/aws](https://twitch.tv/aws)

# Thank you!

