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

AIM306-R

Automate content moderation and compliance with AI

Nate Bachmeier

Sr. Solutions Architect
AWS

John Rouse

Sr. Business Development Lead
AWS

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

Agenda



1. The user-generated content era
2. Moderation capabilities and demo
3. Design patterns, customer stories, and solutions
4. How to get started
5. Q&A

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

The user-generated content (UGC) era

MODERATION COSTS, RISK, AND LIABILITY FROM HARMFUL EXPOSURE ON THE RISE

26% UGC platform industry growth ($20 billion by 2028)[1]

79% of consumer decisions are influenced by UGC[2]

80% of all web content is UGC[3]

[1] Grand View Research, Market Analysis Report (bit.ly/3Evj6r7)
[2] Nosto, Customer Survey (bit.ly/3i7jLHL)
[3] Gorilla Group Consultants, USG Insights (bit.ly/3GEyuEc)

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



Moderation scale, speed, and accuracy at a lower cost

Needs

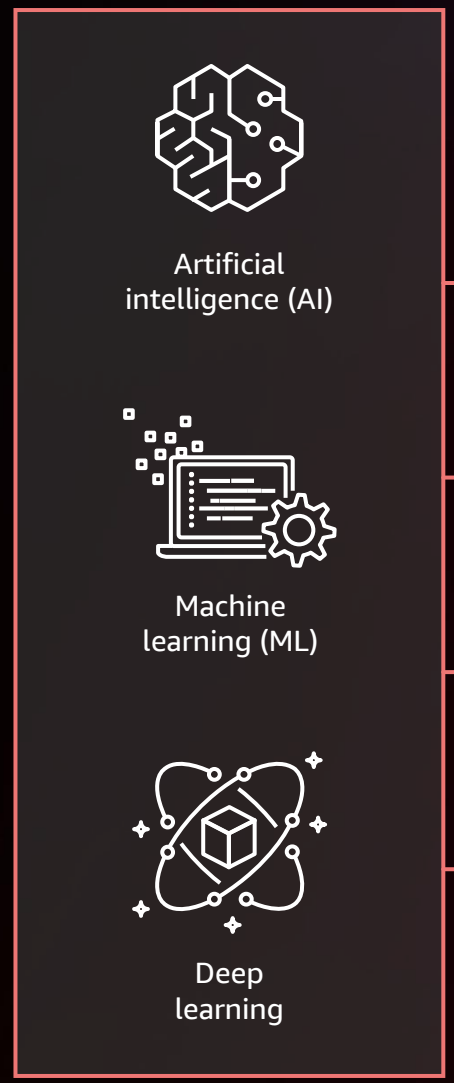
Moderate high volumes of UGC

Keep online communities safe and inclusive

Better allocate human skill

Meet compliance and guidelines

Reduce moderation costs



Benefits

Improve brand, information, and user safety

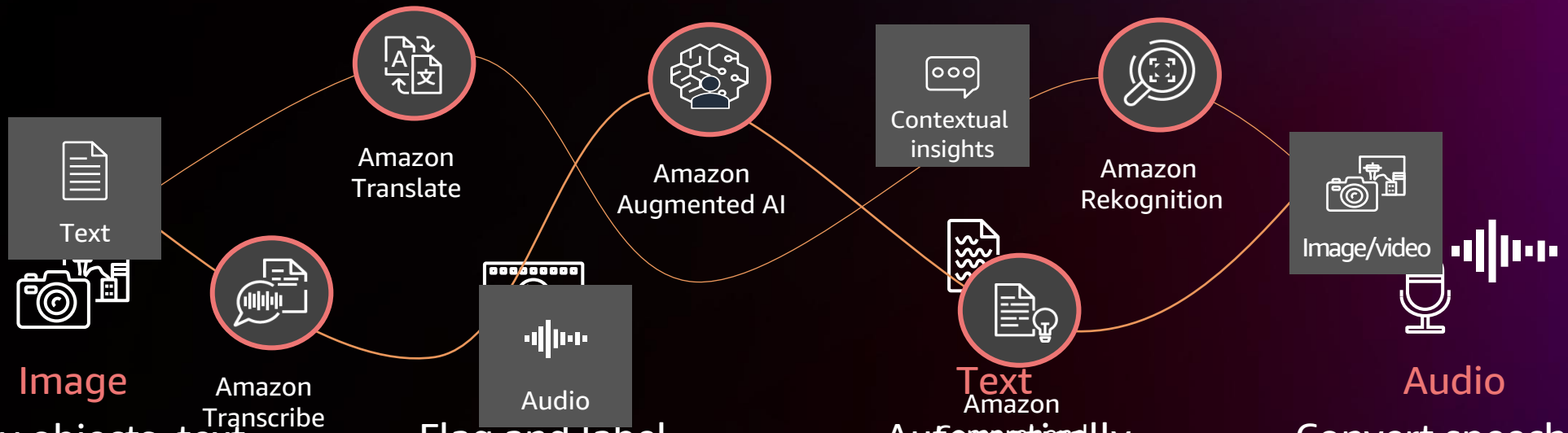
Scale coverage

Reduce human moderators' workload

Save time and costs

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

Multimodal capabilities with AWS AI and ML



Image

Identify objects, text, logos, scenes, and filter harmful visuals and unwanted associations

Flag and label inappropriate content across hierarchical categories

Text

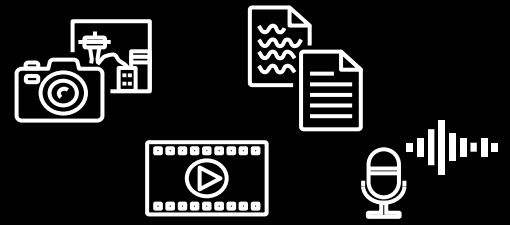
Automatically identify profanity, filter, and redact personal information

Audio

Convert speech to text to filter profanity and any unwanted vocabulary

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

Lower cost and speed to market with AI across industries



Multimodal moderation

Amazon Rekognition

Amazon Transcribe

Amazon Translate

Amazon Comprehend

AI insights and connections

Amazon Augmented AI

Human reviews

Social media

Gaming

E-commerce

Advertising

Highly regulated industries

FinServ

Health

Education

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

Customers expect inclusion and safety

50% of people create content at least once daily[1]

40% of consumers disengage after one exposure to toxic UGC[2]

70% of consumers agree that brands are responsible to moderate UGC[2]



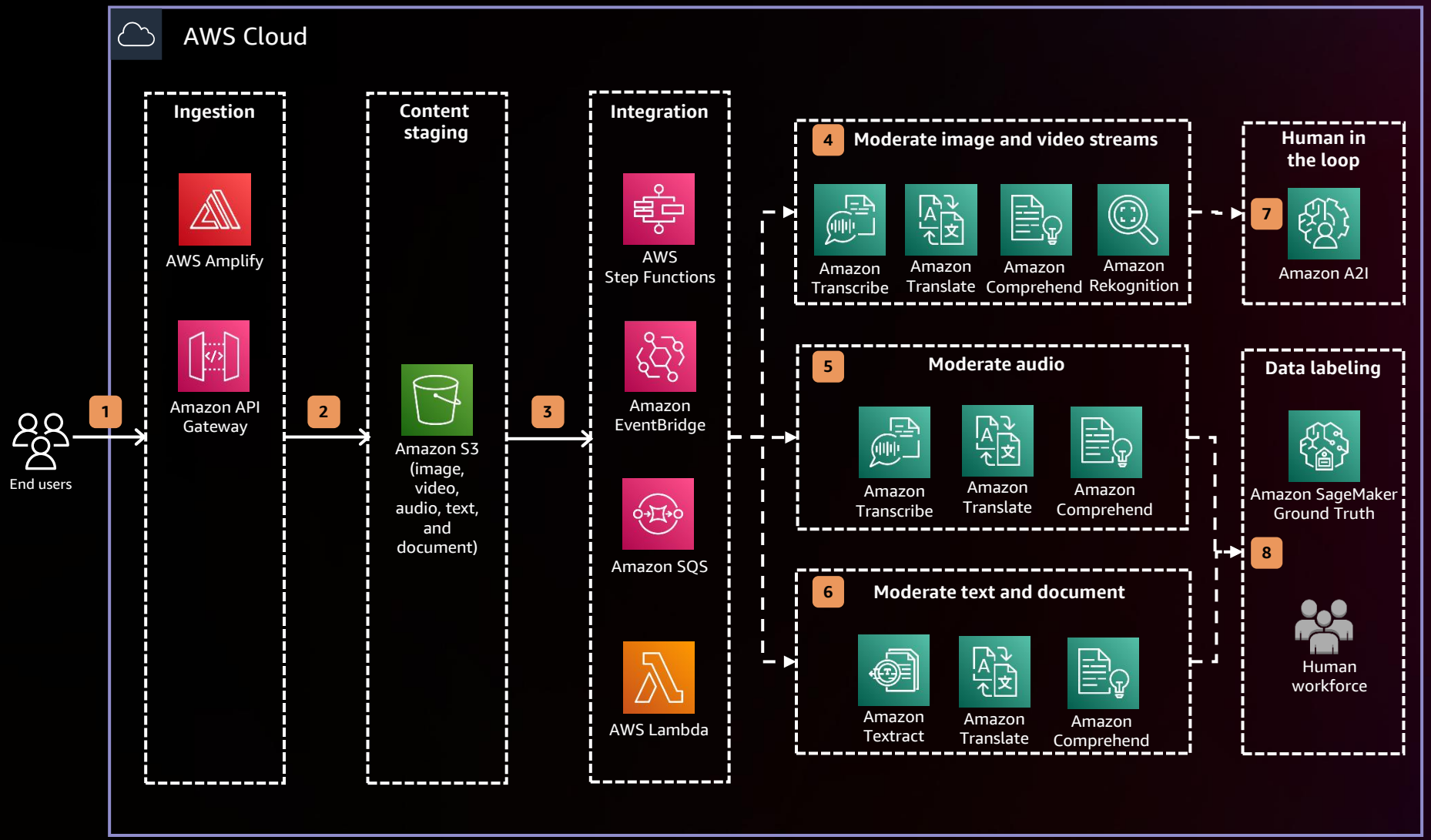
[1] Nosto, Customer Survey (bit.ly/3i7jLHL)
[2] BusinessWire, Customer Survey (bwnews.pr/3AB33GT)

aws

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

Bringing all AWS AI services together

New! AWS Guidance

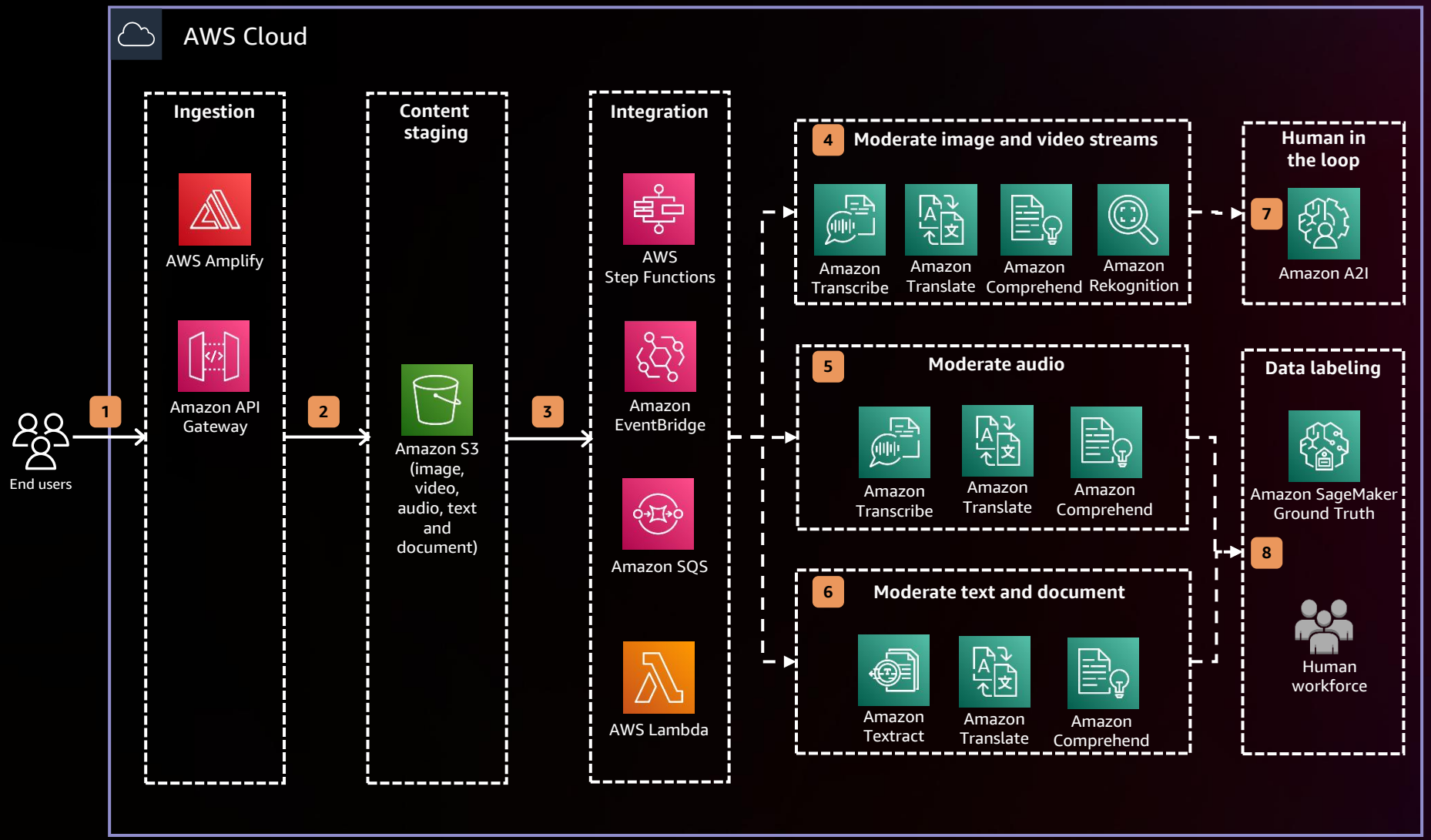


1. Customers upload their content into the **AWS Cloud**
2. Content securely persists into an **Amazon Simple Storage Service (Amazon S3)** bucket or another data store
3. Activate workflows, publisher/subscription patterns, and custom code to moderate content
4. Process the audio streams within video streams using **Amazon Transcribe** and **Amazon Rekognition** and extract content moderation categories using simple APIs
5. Use **Amazon Transcribe** to convert audio into text and leverage natural language processing (NLP) with **Amazon Comprehend**
6. Use **Amazon Textract** to extract content and **Amazon Comprehend** natural language processing to moderate content
7. Bring humans into the loop for scenarios that are not fully automatable using **Amazon Augmented AI (Amazon A2I)**
8. Integrate human workforces to customize model vocabularies and image labels using **Amazon SageMaker Ground Truth**

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

Bringing all AWS AI services together

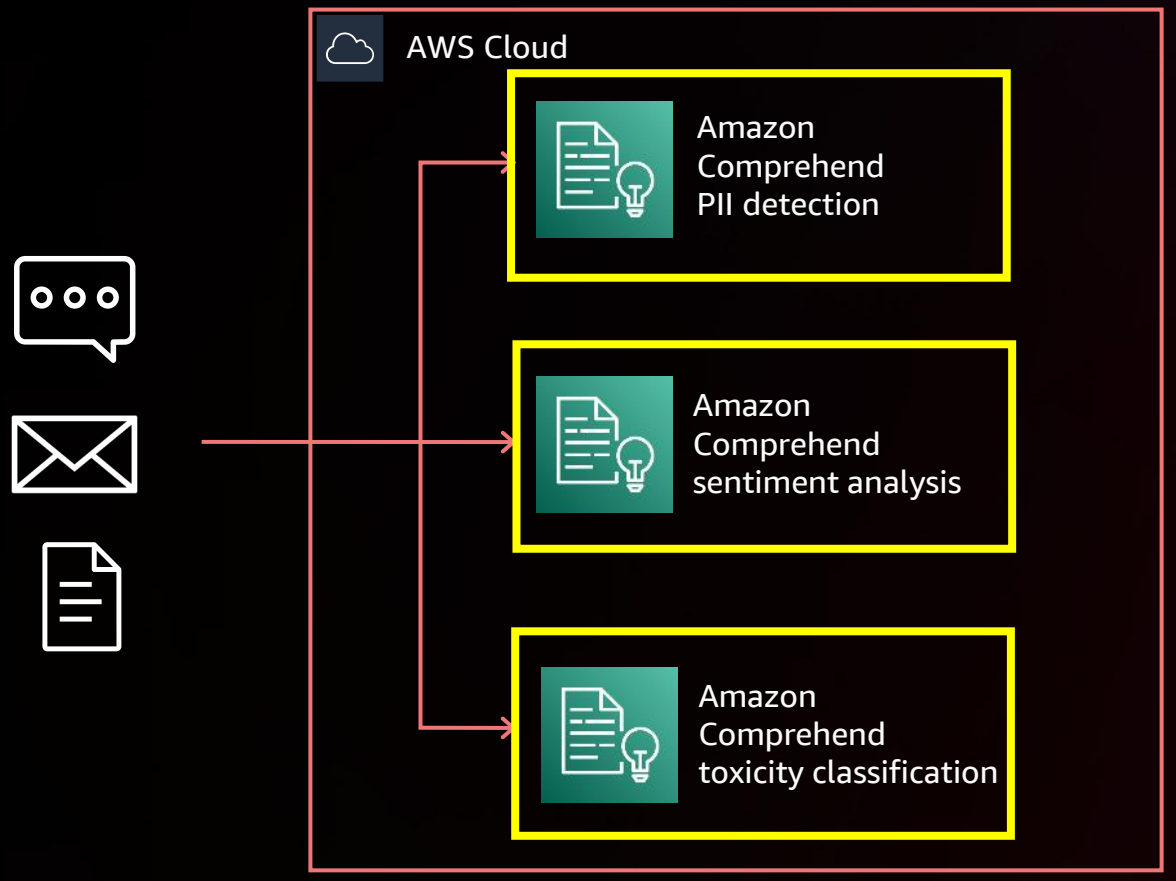
New! AWS Guidance



1. Customers upload their content into the **AWS Cloud**
2. Content securely persists into an **Amazon Simple Storage Service (Amazon S3)** bucket or another data store
3. Activate workflows, publisher/subscription patterns, and custom code to moderate content
4. Process the audio streams within video streams using **Amazon Transcribe** and **Amazon Rekognition** and extract content moderation categories using simple APIs
5. Use **Amazon Transcribe** to convert audio into text and leverage natural language processing (NLP) with **Amazon Comprehend**
6. Use **Amazon Textract to** extract content and **Amazon Comprehend** natural language processing to moderate content
7. Bring humans into the loop for scenarios that are not fully automatable using **Amazon Augmented AI (Amazon A2I)**
8. Integrate human workforces to customize model vocabularies and image labels using **Amazon SageMaker Ground Truth**

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

Uncovering connections in text



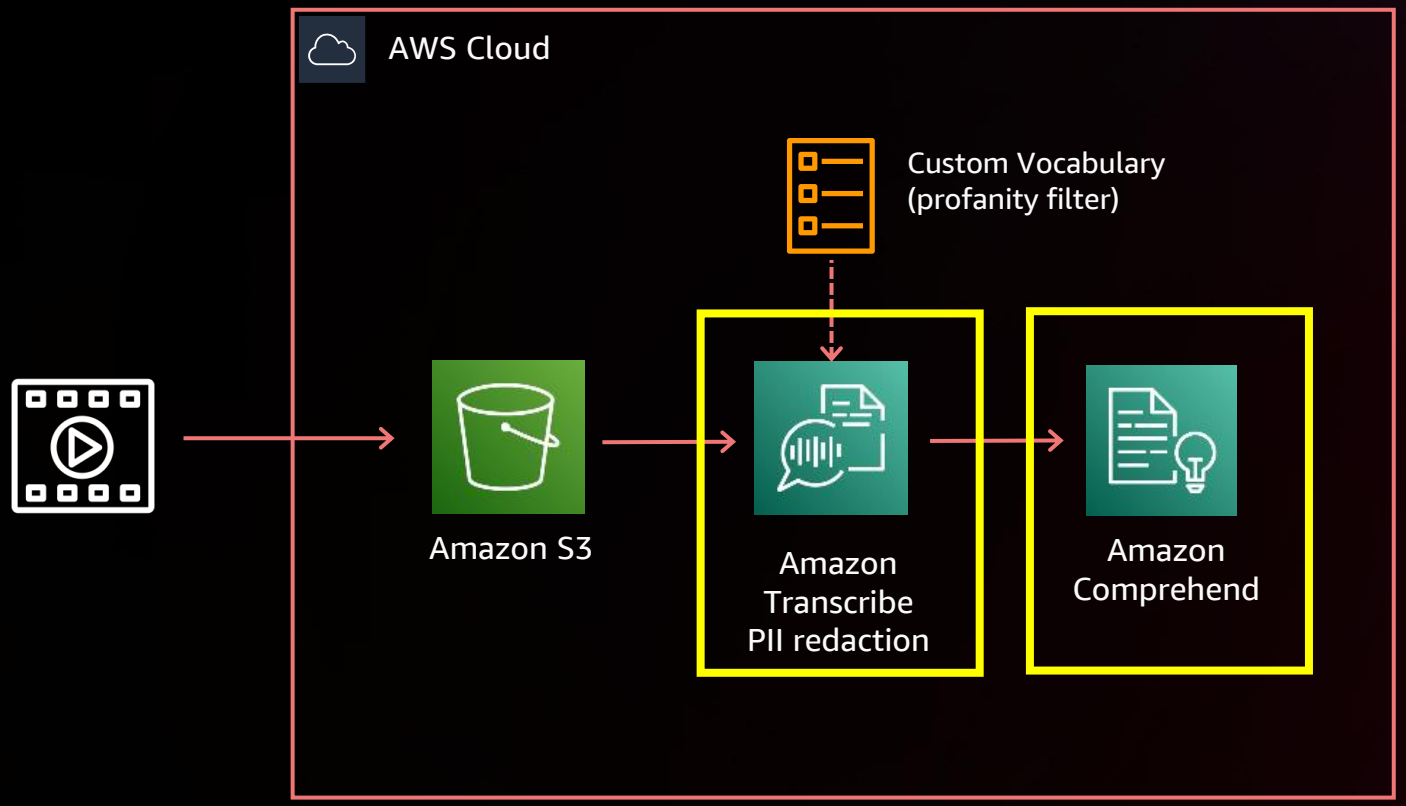
Amazon Comprehend
DetectPiiEntity API
Detect and redact PII entities

Amazon Comprehend
DetectSentiment API
Inspect text sentiment

Amazon Comprehend
Custom Classifier
Classify toxicity sentences

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

Moderate audio of the stored media files



Amazon Transcribe
StartTranscriptionJob API
Transcribing audio to text with
vocabulary filter
and PII redaction

Amazon Comprehend
for text moderation

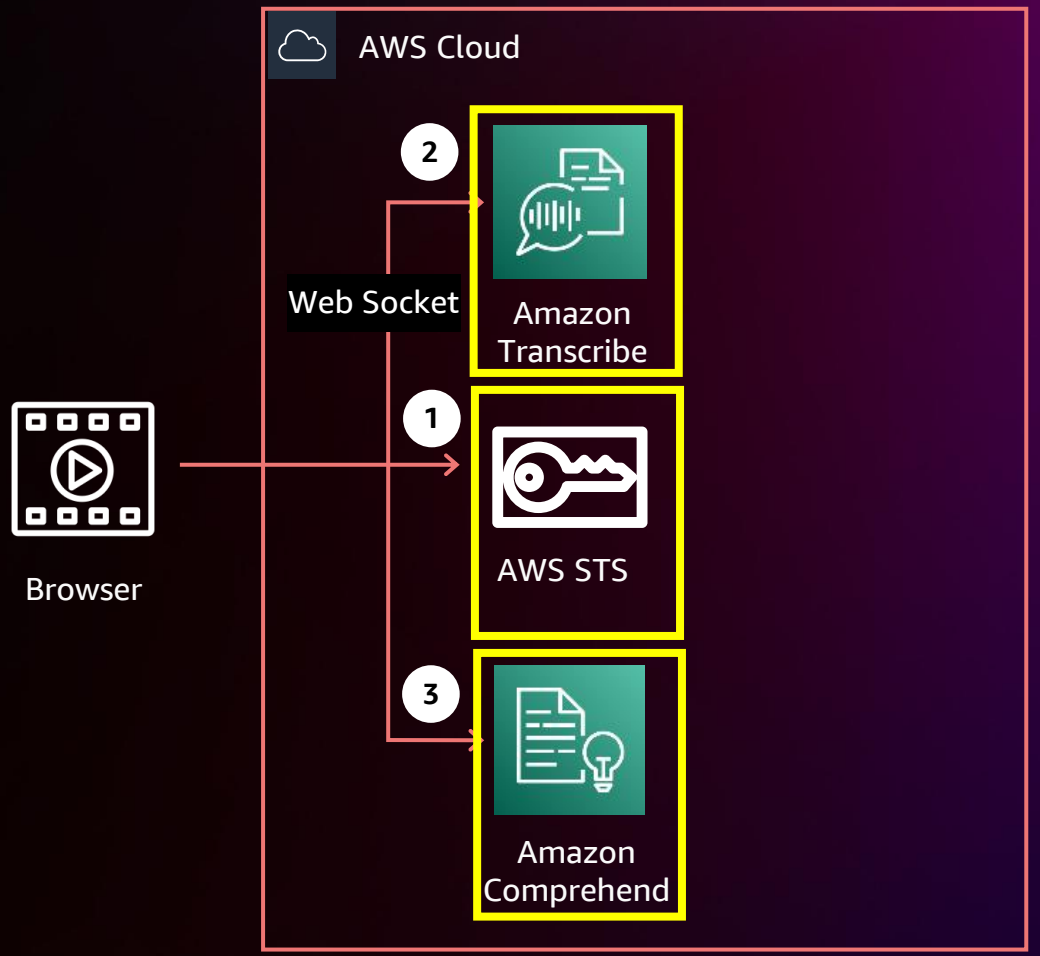
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Live audio moderation

Get STS token

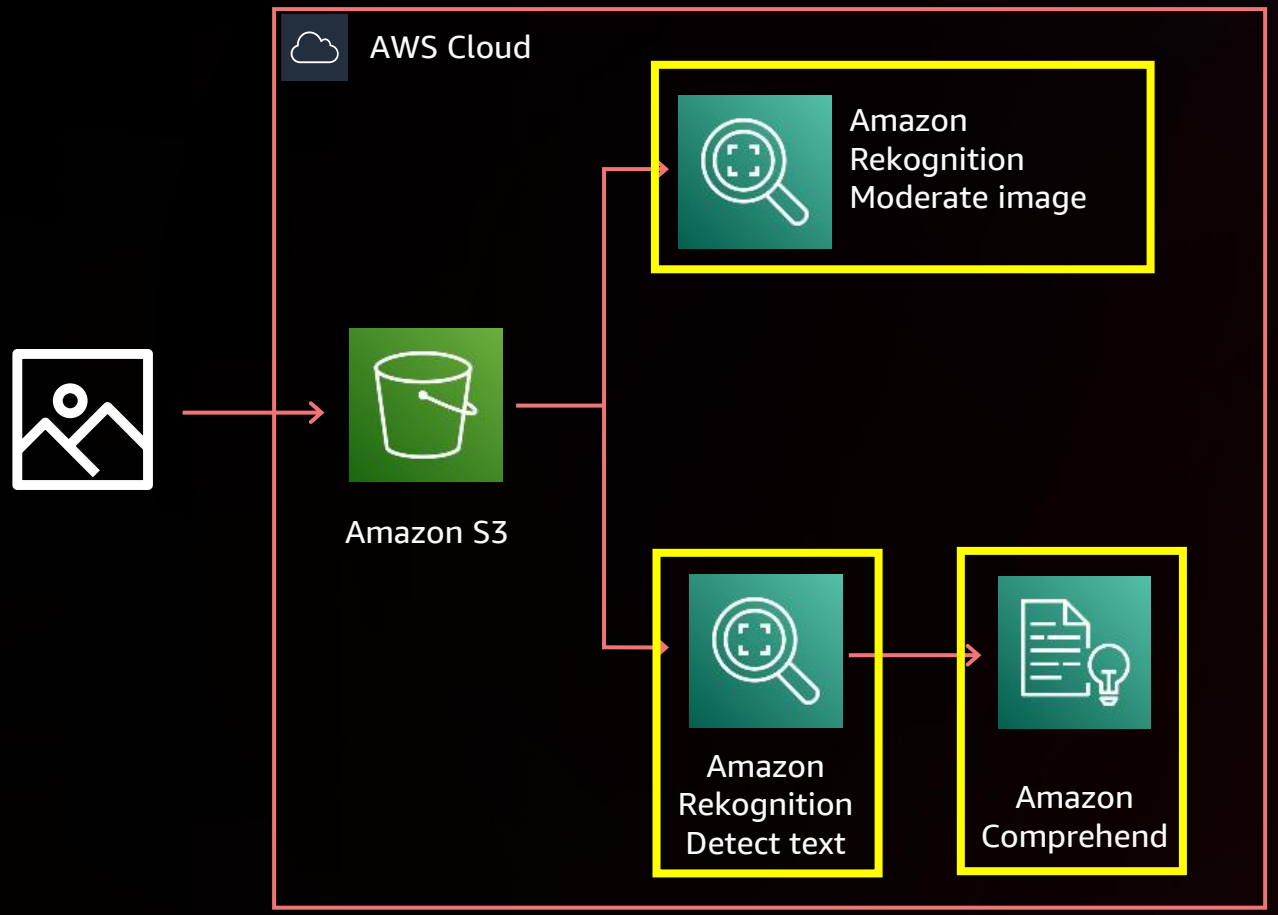
Amazon Transcribe
real-time streaming API
Transcribing audio to text

Amazon Comprehend
For further text moderation



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

Images moderation



Amazon Rekognition
DetectModerationLabels API
Detecting unsafe content in images

Amazon Rekognition
DetectText API
Detecting text in images

Amazon Comprehend
For further text moderation

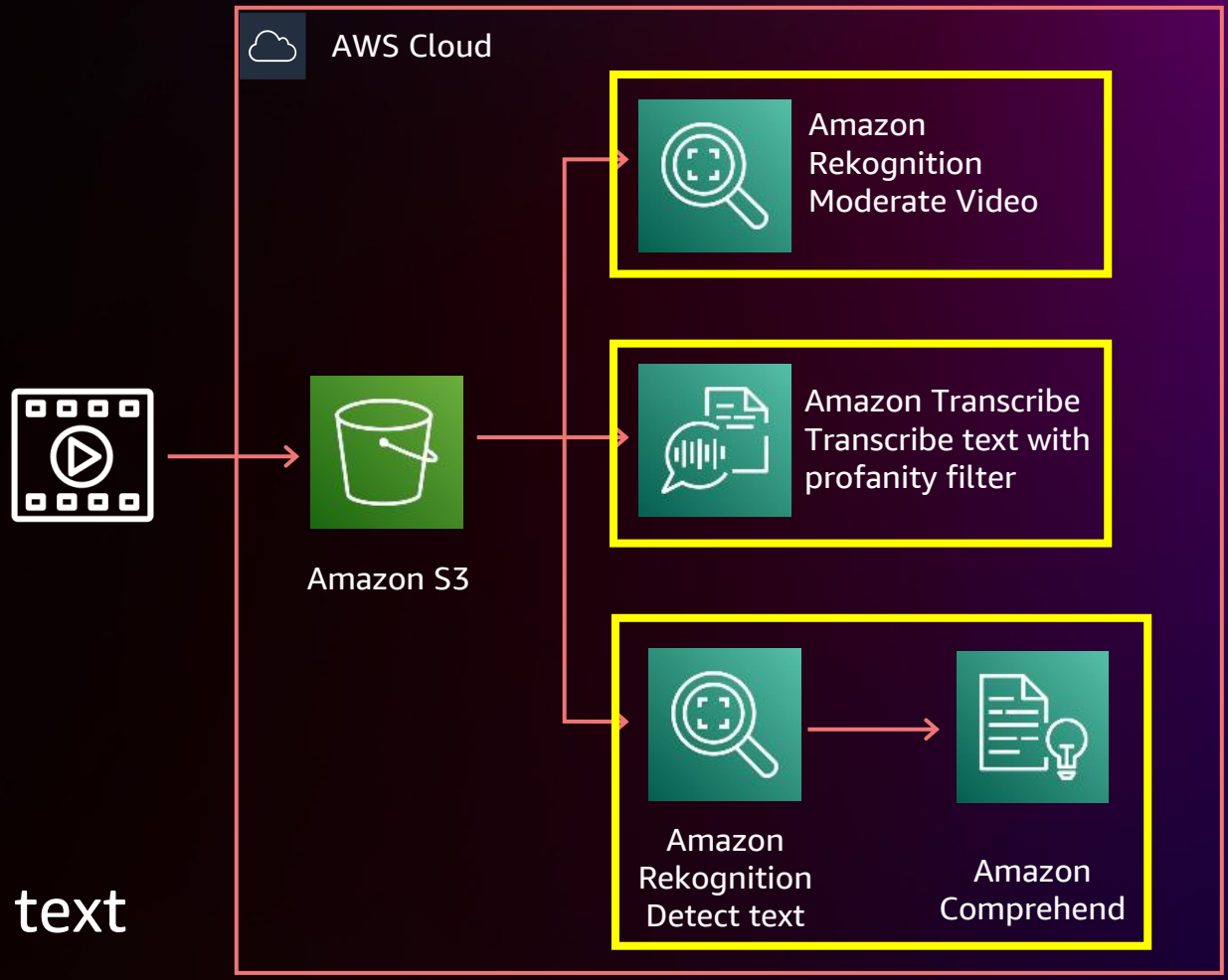
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Moderate stored videos

Amazon Rekognition
DetectModerationLabel API
Detecting unsafe content in video

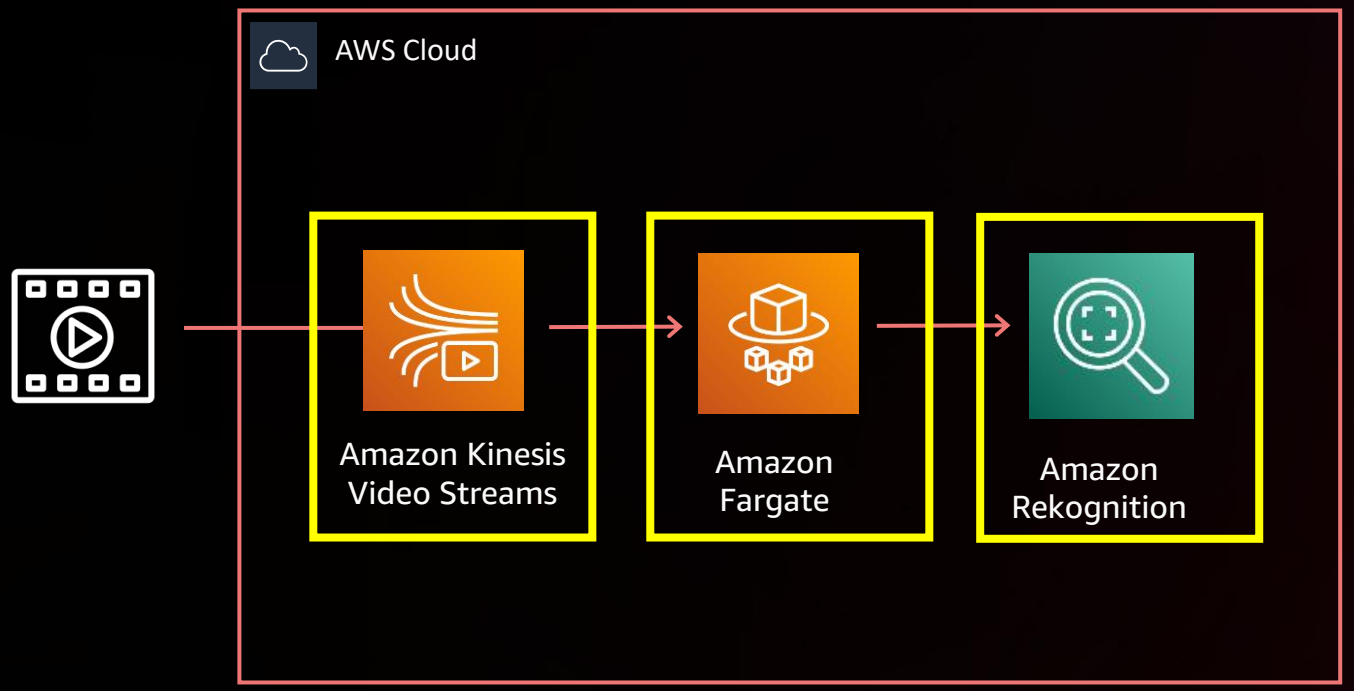
Amazon Transcribe
StartTranscriptionJob API
Transcribing audio to text with custom vocabulary filter

Amazon Rekognition
DetectText API
Detecting text in video and moderate text with Amazon Comprehend



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

Moderate real-time video stream



Amazon Kinesis Video Stream
Consume video streams in real time

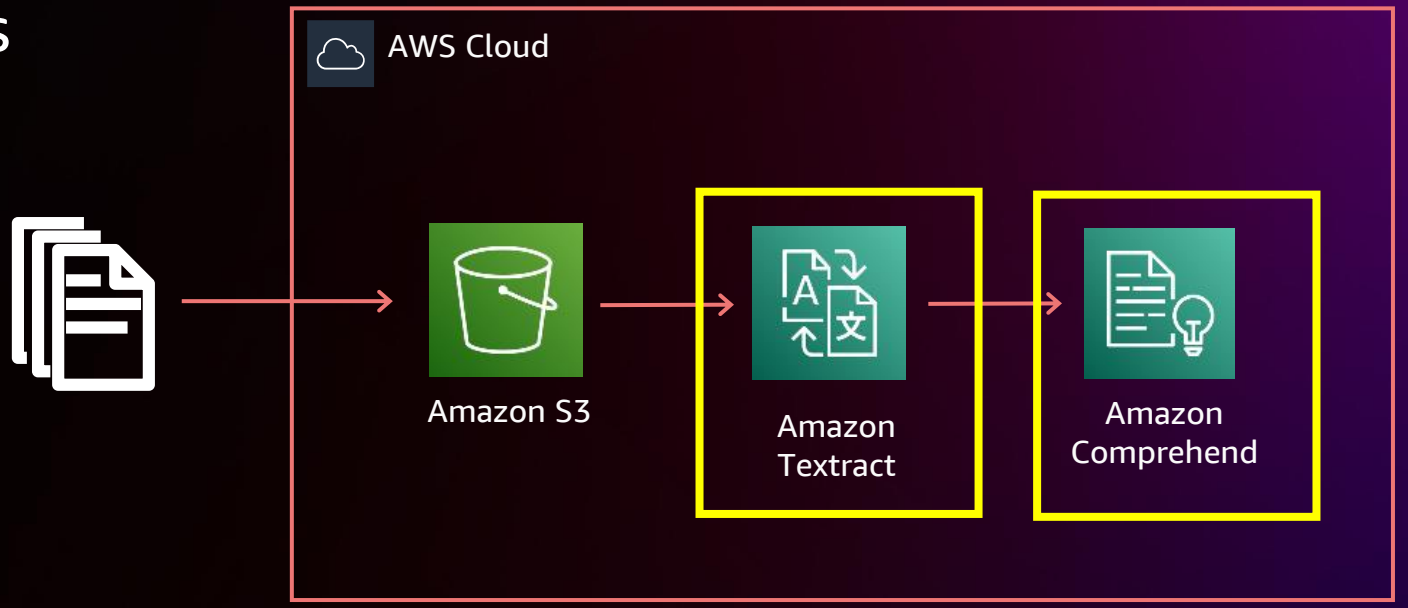
AWS Fargate
Send decoded frames to Amazon Rekognition

Amazon Rekognition
DetectModerationLabels API
Detecting unsafe content in video frames

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

Document moderation

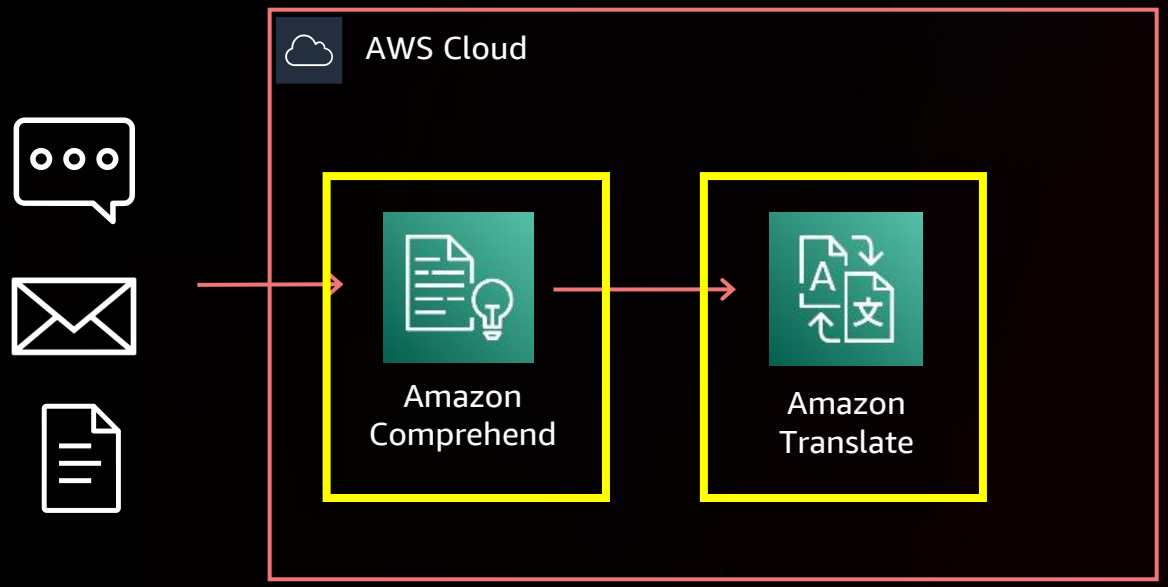
StartDocumentAnalysis API starts the asynchronous analysis of an input document for relationships between detected items



GetDocumentAnalysis API gets the results for an Amazon Textract asynchronous operation that analyzes text in a document

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

Moderation with Amazon Translate



Amazon Comprehend
DetectDomainLanguages API
Detect the language

Amazon Translate
TranslateText API
Translate the text to the target language

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

Highly regulated industries



FinServ



Health



Education

Protection of sensitive information

FinServ	Health	Education
Financial records	Medical records	Media contributions
Applications	Transcriptions	Chats
Tax documents	Protected health information (PHI)	Assignments
Contact center transcripts	Research information	Class transcriptions

Why

International, federal, and state compliance

FinServ	Health	Education
RFPA, GLBA, FCRA, GDPR	HIPAA	FERPA, COPPA, CIPA
Reduce workforce exposure	Secure flow of health information	Foster and increase inclusiveness
	Preserve patients' and subjects' rights	Eliminate toxic environments

aws

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

Customer success stories

:DeNA	FLIPBOARD	MOBISOCIAL
Pococha voice community	Content recommendation platform	Gaming and streaming platform
Identification and redaction of words not recognized by standard speech-to-text services	Automatically detect content violations in 1.2 billion images and videos yearly	Flagging of non-compliant images and videos
Higher detection accuracy and faster moderation response times	Lower false-positive rate and daily interception of 63,000 images and videos	Manual moderation reduced by 95% with increased accuracy and scalability

aws

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

Get started



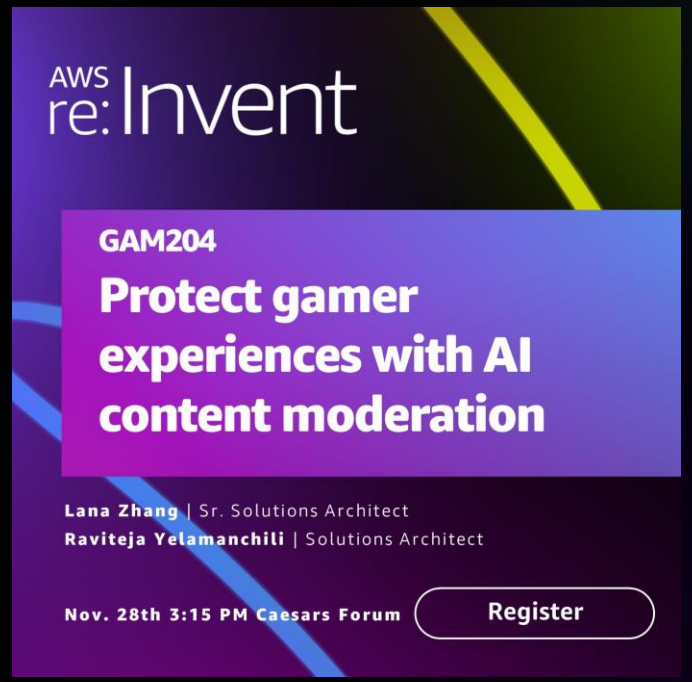
1 **Try and learn**
AWS Content Moderation webpage
New! AWS Guidance, links to design pattern blogs, and more

2 **Build with us**
Contact us

3 **Find a partner**
AWS Partner Network

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

Additional content moderation sessions







Repeats Nov. 30th 12:15 PM
MGM Grand

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

Thank you!

Nate Bachmeier
nbachmei@amazon.com
LinkedIn: nate-bachmeier-b930814a

John Rouse
johnrou@amazon.com
LinkedIn: rousejohn



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

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