Streamline moderation workflows and increase accuracy with Amazon Rekognition Content Moderation version 6

Kevin Carlson
Principal Specialist AWS AI/ML
AWS

Shipra Kanoria
Principal Product Manager Amazon Rekognition
AWS

Agenda

1. The user-generated content era
2. Customer expectations in the UGC era
3. Capabilities and demo of Amazon Rekognition Content Moderation version 6
4. Customer stories, and solutions
5. How to get started

The user-generated content (UGC) era

Moderation costs, risk, and liability from **harmful exposure on the rise**

80% of all web content is UGC[3]

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

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

[1] Grand View Research, Market Analysis Report
[2] Stackla, Customer Survey
[3] Gorilla Group Consultants, USG Insights

Moderation scale, speed, and accuracy at a lower cost

Needs

Moderate high volumes of UGC

Keep online communities safe and inclusive
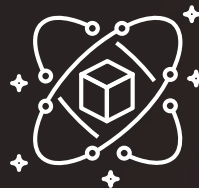
Better allocate human skill

Meet compliance and guidelines

Reduce moderation costs

Artificial intelligence (AI)

Machine learning (ML)

Deep learning

Benefits

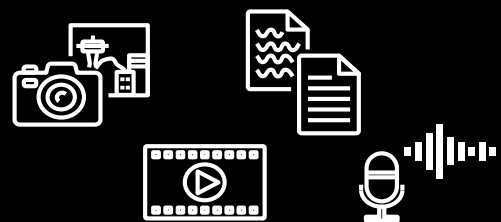Improve brand, information, and user safety

Scale coverage

Reduce human moderators' workload

Save time and costs

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

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]
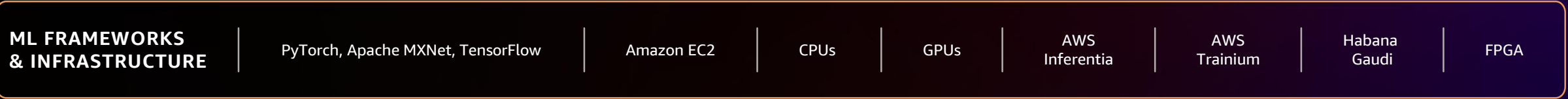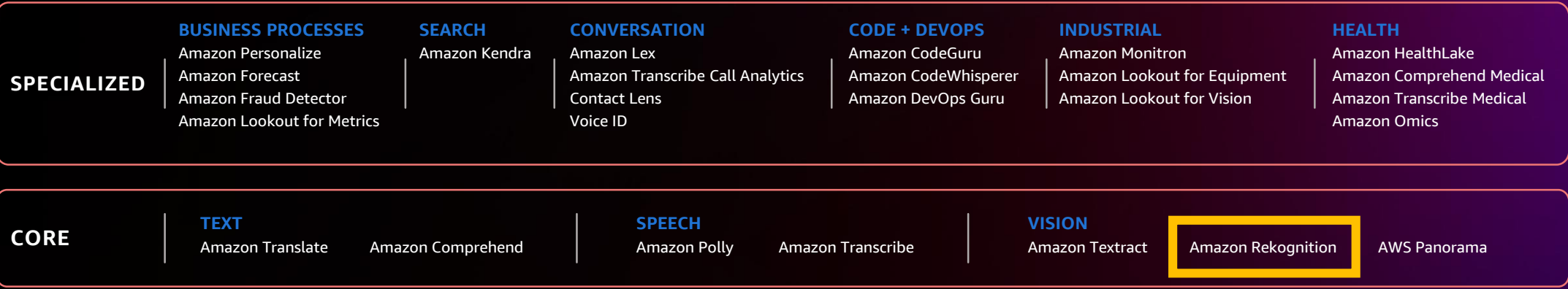
[1] Stackla, Customer Survey
[2] BusinessWire, Customer Survey

The AWS AI/ML stack
Broadest and most complete set of machine learning capabilities

AI SERVICES

SPECIALIZED

BUSINESS PROCESSES
Amazon Personalize
Amazon Forecast
Amazon Fraud Detector
Amazon Lookout for Metrics

SEARCH
Amazon Kendra

CONVERSATION
Amazon Lex
Amazon Transcribe Call Analytics
Contact Lens
Voice ID

CODE + DEVOPS
Amazon CodeGuru
Amazon CodeWhisperer
Amazon DevOps Guru

INDUSTRIAL
Amazon Monitron
Amazon Lookout for Equipment
Amazon Lookout for Vision

HEALTH
Amazon HealthLake
Amazon Comprehend Medical
Amazon Transcribe Medical
Amazon Omics

CORE

TEXT
Amazon Translate   Amazon Comprehend

SPEECH
Amazon Polly   Amazon Transcribe

VISION
Amazon Textract   Amazon Rekognition   AWS Panorama

AMAZON SAGEMAKER

CANVAS
No-code ML for business analysts

STUDIO LAB
Learn ML

GROUND TRUTH
Label data

STUDIO IDE

Prepare data

Geospatial ML

Build with notebooks

Train models

Deploy in production

Store features

Tune parameters

Manage and monitor

-------------------------------- CI/CD | GOVERNANCE | RESPONSIBLE ML --------------------------------

EDGE MANAGER
Manage edge devices

ML FRAMEWORKS & INFRASTRUCTURE

PyTorch, Apache MXNet, TensorFlow | Amazon EC2 | CPUs | GPUs | AWS Inferentia | AWS Trainium | Habana Gaudi | FPGA

aws

Amazon Rekognition for computer vision success

Benefits

Capabilities

Build faster

Reduce operational cost with ready to use resources that reduce operational time implementation, learning and overall processing.

Simplify operations

Fully managed APIs help customers train and implement custom models with less effort and data.

Customize easily

Ready to use models and services that enable developer to quickly implement content moderation workflows and let the service do the heavy lifting.

Amazon Rekognition
Automated image, video, and text analysis

CONTENT MODERATION

IDENTITY VERIFICATION

FACE DETECTION, ANALYSIS, COMPARE, AND SEARCH
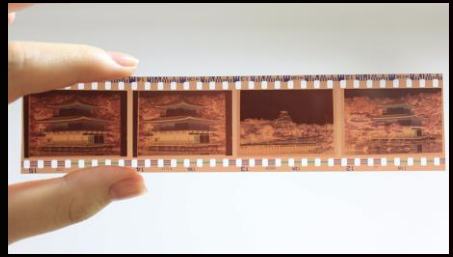
LABELS

CUSTOM LABELS

TEXT DETECTION

CELEBRITY RECOGNITION

VIDEO SEGMENT DETECTION

STREAMING VIDEO EVENTS DETECTION

Content Moderation

Identity verification

Connected home

Media analysis & compliance

aws

Content moderation with AWS

Automate **image, video, text, and audio** moderation workflows with fully automated AI services to moderate unwanted content and protect sensitive information

Amazon Rekognition	Amazon Transcribe	Amazon Translate	Amazon Comprehend	Amazon Augmented AI
Image and video analysis	Speech to text	Translation	Natural language processing (NLP)	Human reviews

Multimodal capabilities with AWS AI and ML

AMAZON TRANSLATE

CONTEXTUAL INSIGHTS

AMAZON AUGMENTED AI

AMAZON REKOGNITION

TEXT

IMAGE

AMAZON TRANSCRIBE

AUDIO

TEXT
AMAZON COMPREHEND

IMAGE/VIDEO

AUDIO

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

Content moderation API

Top-level category	Second-level category
Explicit Nudity	Nudity
	Graphic Male Nudity
	Graphic Female Nudity
	Sexual Activity
	Illustrated Explicit Nudity
	Adult Toys
Suggestive	Female Swimwear or Underwear
	Male Swimwear or Underwear
	Partial Nudity
	Barechested Male
	Revealing Clothes
	Sexual Situations
Violence	Graphic Violence or Gore
	Physical Violence
	Weapon Violence
	Weapons
	Self-Injury
Visually Disturbing	Emaciated Bodies
	Corpses
	Hanging
	Air Crash
	Explosions and Blasts

Top-level category	Second-level category
Rude Gestures	Middle Finger
Drugs	Drug Products
	Drug Use
	Pills
	Drug Paraphernalia
Tobacco	Tobacco Products
	Smoking
Alcohol	Drinking
	Alcoholic Beverages
Gambling	Gambling
Hate Symbols	Nazi Party
	White Supremacy
	Extremist

Detect inappropriate content in images and videos across 10 categories and 35 sub-categories

Get granular output about the nature of the content, rather than basic safe/unsafe flagging.

Apply your own policies that define what is safe for your customers.

aws

Amazon Rekognition

Improvements in V6 (Image API only)

- Improved false detection (FPR) for selfie images

- Improved false detection (FPR) across many other categories such as

  household objects, clothing items and accessories

Amazon Rekognition

Why these improvements are important

Lower false detections

Lower volume of images to be reviewed

Less cost

Content moderation API

Top-level category	Second-level category
Explicit Nudity	Nudity Graphic Male Nudity Graphic Female Nudity Sexual Activity Illustrated Explicit Nudity Adult Toys
Suggestive	Female Swimwear or Underwear Male Swimwear or Underwear Partial Nudity Barechested Male Revealing Clothes Sexual Situations
Violence	Graphic Violence or Gore Physical Violence Weapon Violence Weapons Self-Injury
Visually Disturbing	Emaciated Bodies Corpses Hanging Air Crash Explosions and Blasts

Top-level category	Second-level category
Rude Gestures	Middle Finger
Drugs	Drug Products Drug Use Pills Drug Paraphernalia
Tobacco	Tobacco Products Smoking
Alcohol	Drinking Alcoholic Beverages
Gambling	Gambling
Hate Symbols	Nazi Party White Supremacy Extremist

Detect inappropriate content in images and videos across 10 categories and 35 sub-categories

Get granular output about the nature of the content, rather than basic safe/unsafe flagging.

Apply your own policies that define what is safe for your customers.

aws

Demo

Marketplaces: Screen product images to detect inappropriate or prohibited product listings

An image of a purse

Moderation API Response

```
{
"model":"Content Moderation V6.0",
"results":[]
}
```

Demo

Marketplaces: Screen product images to detect inappropriate or prohibited product listings

A purse with a partially-visible handgun

Moderation API Response

```
{
"model":"Content Moderation V6.0",
"results":[
{
"name":"Weapon Violence",
"parent":"Violence",
"score":90.579343
},
]
}
```

Demo

Marketplaces: Screen product images to detect inappropriate or prohibited product listings

An image of a watch

Moderation API Response

```
{
"model":"Content Moderation V6.0",
"results":[]
}
```

Demo

Marketplaces: Screen product images to detect inappropriate or prohibited product listings

Visible Male Barechest

Moderation API Response

```
{
"model":"Content Moderation V6.0",
"results":[
{
"name":"Barechested Male",
"parent":"Suggestive",
"score":95.988031
},

]
}
```

Demo

Online Communities & Social Media: Identify inappropriate content in profile images and user-uploaded images and videos to create a safe online community

An appropriate profile image

Moderation API Response

```
{
"model":"Content Moderation V6.0",
"results":[]
}
```

Demo

Online Communities & Social Media: Identify inappropriate content in profile images and user-uploaded images and videos to create a safe online community

Prohibited profile image

Moderation API Response

```
{
"model":"Content Moderation V6.0",
"results":[
{
"name": "Smoking",
"parentName": "Tobacco"
"score":95.17668
},
]
}
```

Demo

Online Communities & Social Media: Identify inappropriate content in profile images and user-uploaded images and videos to create a safe online community

Prohibited profile image

Moderation API Response

```
{
"model":"Content Moderation V6.0",
"results":[
{
"name":"Drinking",
"parent":"Alcohol",
"score":91.384314
},
]
}
```

Demo

Online Communities & Social Media: Identify inappropriate content in profile images and user-uploaded images and videos to create a safe online community

Prohibited profile image

Moderation API Response

```
{
"model":"Content Moderation V6.0",
"results":[
{
"name":"Middle Finger",
"parent":"Rude Gestures",
"score":62.775986
},
]
}
```

Demo

Marketing & Advertising: Detect inappropriate brand imagery in campaigns and creatives

Appropriate brand imagery

Moderation API Response

```
{
"model":"Content Moderation V6.0",
"results":[]
}
```

Demo

Marketing & Advertising: Detect inappropriate brand imagery in campaigns and creatives

Inappropriate brand imagery

Moderation API Response

```
{
"model":"Content Moderation V6.0",
"results":[
{
"name": "Tobacco Products",
"parent": "Tobacco"
"score":99.696,
},
]
}
```

Demo

Marketing & Advertising: Detect inappropriate brand imagery in campaigns and creatives

Appropriate brand imagery

Moderation API Response

```
{
"model":"Content Moderation V6.0",
"results":[]
}
```

Demo

Marketing & Advertising: Detect inappropriate brand imagery in campaigns and creatives

Inappropriate brand imagery

Moderation API Response

```
{
"model":"Content Moderation V6.0",
"results":[
{
"name":"Weapons",
"parent":"Violence",
"score":90.434961
},
]
}
```

Customer Success Stories

:DeNA

Pococha voice community

Identification and redaction of words not recognized by standard speech-to-text services

Higher detection accuracy and faster moderation response times

FLIPBOARD

Content recommendation platform

Automatically detect content violations in 1.2 billion images and videos yearly

Lower false-positive rate and daily interception of 63,000 images and videos

MOBISOCIAL

Gaming and streaming platform

Flagging of non-compliant images and videos

Manual moderation reduced by 95% with increased accuracy and scalability

Moderation in Action

Get started

1 **Try and learn**
Amazon Rekognition
Thousands of images and videos free per month
for 12 months with the AWS Free Tier

AWS Content Moderation webpage
New! AWS Guidance, links to design pattern blogs, and more
aws.amazon.com/rekognition

2 **Build with us**
Contact us

3 **Find a partner**
AWS Partner Network

Thank you!

Kevin Carlson

kvcarls@amazon.com

Shipra Kanoria

skanoria@amazon.com