



**Protect your users,
brand, and budget
with AI-powered
content moderation**



Table of contents

Introduction.....	3
A surge in user-generated content.....	4
Human-based content moderation is getting overwhelmed	5
AWS delivers automated, intelligent content moderation	6
How it works.....	10
Get started with AWS AI services for content moderation	16

Introduction

As the volume, complexity, and speed of user-generated content (UGC) increase, organizations of all sizes must commit significant technical and human resources to ensure customers are not exposed to offensive material. At the same time, they're under pressure to take advantage of new opportunities and stand out from the competition—all while keeping operational costs from spiraling out of control. Human-based content moderation alone cannot scale to meet the safety, regulatory, and operational needs of the UGC era, leading to poor user experience, high moderation costs, and brand risk. Artificial Intelligence and Machine Learning (AI/ML) can help organizations achieve their objectives while reducing costs and human moderation workloads. AWS has AI solutions and services that can be self-managed or customized to help you create safe online communications and protect your customers and brand, all while reducing moderation costs.



A surge in user-generated content

The daily volume of UGC is increasing exponentially across every platform that invites users to post comments, audio, documents, images, and videos, as well as to chat or talk live with others. A staggering 80 percent of all web content is now UGC.¹ And while this provides organizations with incredible opportunities to create new ways to connect with their audiences, it also increases the requirements and costs of keeping content and online communities safe and compliant. 79 percent of consumers report that their purchasing decisions are being influenced by UGC,² and 40 percent say they would disengage with a brand if exposed to toxic or fake UGC on its channels. Seventy percent believe that brands need to protect users from toxic content and 78 percent say it is a brand's responsibility to provide positive and welcoming online experiences.³ These trends are putting pressure on organizations to protect their customers while reducing costs.

¹ Harrison, V., "User-Generated Content Is the New World-of-Mouth," Gorilla Group, May 2022

² "43 Statistics About User-Generated Content You Need to Know," Nosto Solutions Oy, April 2022

³ "Nearly Half of Americans Quickly Lose Trust in a Brand If Exposed to Toxic or Fake User-Generated Content on Its Channels," Businesswire, June 2021



78 percent
say it is a brand's responsibility
to provide positive and welcoming
online experiences.

Human-based content moderation is getting overwhelmed

Today's consumers are creating and consuming massive amounts of UGC as part of their daily online routine. On Facebook alone, 1 billion stories are shared every day.⁴ And today's UGC comes in variety of forms from an increasing number of unique platforms. Take gaming, for example. There are more than 2.7 billion gamers worldwide, and 60 percent of Americans play video games at least once daily,⁵ creating one of the largest platforms for UGC that is growing by the second.

The sheer volume of UGC creates significant challenges for companies of all sizes. Some spend big to employ large teams of human moderators, either in-house or outsourced. Based on volume and complexity, teams can rapidly go from a few members to a few thousands. Others rely on smaller teams and can only react when users have already been exposed to inappropriate content. Big or small, both approaches have consequences:

Big teams are expensive, leading to high moderation costs. Plus, the workloads on those teams can be exhausting, time consuming, and difficult to scale.

Big or small, any moderation team that misses something or reacts after the fact delivers a poor user experience. Users who are exposed to offensive material are likely to leave a site immediately, leaving the risk of brand damage and liability.

Moderation teams of all sizes and workflows need to operate at close to real-time speed in order to avoid users churning to other platforms that provide safer and more inclusive experiences.

As the generation of third-party content continues to rise across all industries, these human-based content moderation models are becoming less sustainable from a financial, operational, and a human level. Organizations are under pressure to protect their customers, brands, and revenue streams from harm and innovate fast to tap into new opportunities and stand out from a growing list of competitors while reducing moderation costs and human workload.

⁴ Bagadiya, J., "430+ Social Media Statistics You Must Know in 2022," SocialPilot Technologies Inc. 2022

⁵ "45+ Out of This World Video Games Industry Statistics in 2022," Techjury 2022

AWS delivers automated, intelligent content moderation

Content moderation is not a one-size-fits-all challenge, as almost all consumers in every industry consume UGC via more than one medium source. It's a multi-modal challenge that requires a solutions-based approach. AWS offers fully managed AI services that can be implemented without ML expertise quickly and easily, tailored to your specific content moderation needs. AWS machine learning services work together to make it easy and scalable to moderate content across a variety of different media types, including audio, video, image, text, speech, and documents. Providing automation and AI capabilities to implement a reliable content moderation solution without requiring ML expertise.

Streamline your content moderation workflows

AWS AI services and solutions are designed to help you create reliable, scalable, and repeatable cloud-based content moderation workflows without the financial burden of upfront commitments or expensive licenses. Your organization can automate tasks completely, or significantly reduce human-in-the-loop moderation with AI/ML predictions.

AWS AI services let you scale coverage, allowing your organization to explore new revenue streams, process a growing influx of UGC on existing channels, continue to grow your online communities, and deliver exceptional user experiences. AI services review every content piece to make sure that nothing slips through the cracks, translate text, and automatically identify and redact unwanted content or personal identifiable information. You can even integrate two or more services together to create a unique blend of multi-modal moderation capabilities to match the needs of your organization.

Organizations are turning to AI and ML for...

Scale: Content moderation is not a one-size-fits-all challenge, as almost all companies in every industry consume UGC in more than one media type.

Speed: AI services automate content moderation tasks, enabling organizations to reclaim up to 95 percent of the time their teams spend doing it manually.

Simplicity: AI self-managed or customizable services help create solutions that can be implemented without AI/ML expertise or tailored to meet specific business needs.

Cost Savings: AI services ensure you get comprehensive yet cost-effective moderation coverage for all your content as your business grows without upfront costs of purchase and installation, as well as ongoing costs like maintenance and upgrades.



SOCIAL MEDIA

Protect brands from unwanted associations and minimizing brand risk



Flipboard uses AWS AI/ML services to automatically detect violations in the nearly 1.2 billion pieces of content uploaded to its platform yearly. On average, the platform finds 63,000 violations per day with lower false-positive rates, more manageable workloads for its moderators, and lower operational costs.

Create safe and enjoyable user experiences

More than 40 percent of respondents will disengage from a brand's community after as little as one exposure to toxic or fake UGC. AWS AI services can help customers feel more involved, engaged, and welcomed by protecting them from unwanted content and behaviors. Index and flag violations across large volumes of content and detect unwanted content in objects, scenes, and custom labels within images.

ECOMMERCE

Boost marketplace value and engagement



11 Street Co continues to empower buyers and sellers to express themselves freely, but within platform guidelines without affecting performance, quality, or user experience. Using AWS AI/ML tools, 11 Street can process more than 7,000 daily images and videos with a higher success rate and lower cost than its previously manual processes.



Reduce moderation costs

Simply pay by the number of images or the duration of videos that you want to process (**Amazon Rekognition**), the total number of characters translated (**Amazon Translate**), seconds of audio transcribed per month with natural language processing (NLP) and speech-to-text capabilities (**Amazon Transcribe**), and scale as needed. And you can extend the value of your investment with additional contextual analysis from **Amazon Comprehend**, and with strategic human-in-the-loop moderation in the loop with **Amazon Augmented AI**.

SOCIAL MEDIA

Automatically review photo libraries of any scale

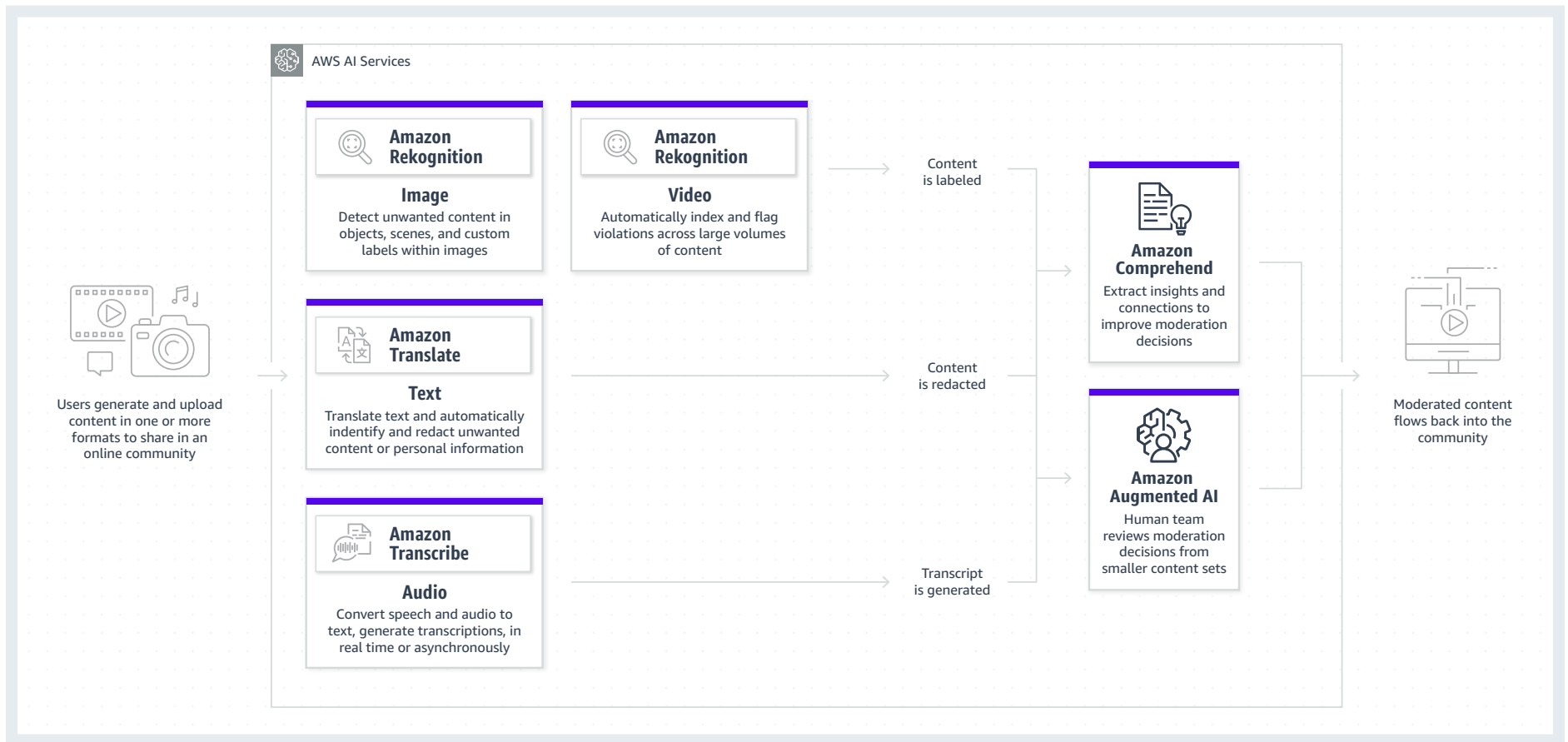


Coffee Meets Bagel reduced moderation costs by 72 percent and human moderator involvement by 97 percent, creating a safe and enjoyable experience for users trying to find their soulmates on its platform and solving moderator workload issues.



How it works

AWS fully managed AI services can work together to provide streamlined content moderation and provide total coverage of all content types. Increase moderation accuracy with additional contextual insights and bring human moderators in the loop to verify decisions when necessary and to improve the ML model over time.



Amazon Rekognition

Amazon Rekognition Content Moderation enables you to streamline or automate your image and video moderation workflows using machine learning. Using fully managed image and video moderation APIs, you can proactively detect inappropriate, unwanted, or offensive content containing nudity, suggestiveness, violence, and other such categories. Amazon Rekognition returns a hierarchical taxonomy of moderation-related labels that make it easy for you to define granular business rules as per your own Standards and Practices (S&P), user safety, or compliance guidelines—without requiring any machine learning experience. You can then use machine predictions to either automate certain moderation tasks completely, or to significantly reduce the review workload of trained human moderators, so that they can focus on higher value work.

REAL ESTATE



Ensure an inclusive, safe data-driven user community

CoStar uses Amazon Rekognition's Content Moderation API to easily build a solution to automatically analyze all uploaded images on its platform. Amazon Rekognition offers a suite of pre-trained computer vision APIs that offer text detection and object detection that help CoStar further improve its product offerings by making the images it receives more discoverable and its community more inclusive.



Amazon Transcribe

Amazon Transcribe enables you to convert speech-to-text, and then check it against custom lists of banned words to detect harmful speech/audio in videos and any other audio-focused content. Transcribe can redact automated live voice transcriptions to ensure profanity is not relayed over subtitles. Transcribe's features enable you to ingest audio input, produce easy-to-read-and-review transcripts, improve accuracy with customization, and filter content to ensure customer privacy.

GAMING AND E-SPORTS

DeNA

Increase member inclusiveness and protection from harmful content

DeNA, one of Japan's leading mobile gaming and internet services companies, now offers a live audio distribution app called Voice Pococha. With Amazon Transcribe and its custom vocabulary feature, DeNA creates a safe community for its users by automatically identifying and redacting banned words not recognized by standard speech-to-text services.



Amazon Translate

Amazon Translate can help you moderate the massive amount of text content out there, including social media comments, reviews, chat rooms, discussions boards, blogs, and documents. Using Custom Terminology, Amazon Translate allows you to define terms or names that are unique to certain organizations, domain, and industry, creating a custom solution that fits your text moderation needs. Translate can redact unwanted content within text, and when enabling the profanity setting, can mask profane words and phrases in your translation output.

ONLINE MARKETPLACE



"By migrating to Amazon Translate, we've been able to implement machine learning models, achieve all the customizations that we need, and get a good quality of service without the complexity."

Franco Yadon
Senior Expert, Mercado Libre



Amazon Comprehend

Amazon Comprehend complements content moderation capabilities by analyzing data and improving the decisions made by core services—Amazon Rekognition, Amazon Translate, and Amazon Transcribe. Content relating to each industry can be very specific. Comprehend learns the unique characteristics of the content being moderated and provides valuable, personalized insights to consistently improve the decision output of your automated content moderation workflow.

SOFTWARE



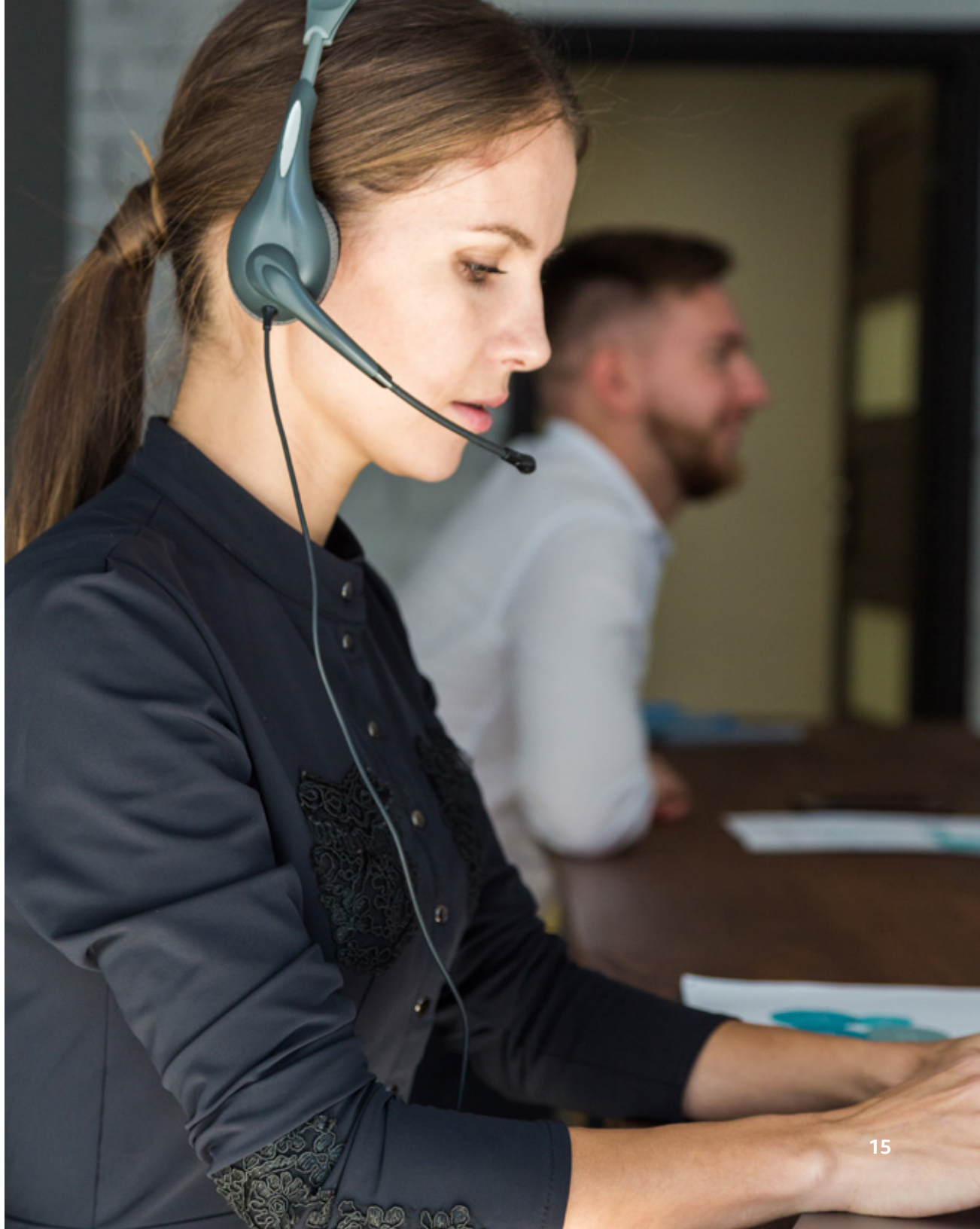
"Amazon Comprehend and Amazon Transcribe services allow VidMob to build high-quality machine learning text analysis into our Agile Creative Suite, enabling us to help brand clients understand content performance in ways never before possible. We are able to transcribe text from video content, and quickly analyze it using Comprehend, allowing us to surface actionable insights to both our creator community and our clients, giving them a strategic edge in the market."

Alex Collmer
Founder and Chief Executive Officer of VidMob



Amazon Augmented AI

Some machine learning applications need human oversight to ensure accuracy with sensitive data, to help provide continuous improvements and retrain models with updated predictions. However, in these situations, companies are often forced to choose between a machine-learning-only or human-only system. **Amazon Augmented AI** provides the best of both worlds with the speed of automated content moderation and the confidence of human review. Empower your human moderation teams by separating high and low-confidence predictions made in your automated workflow and then incorporating human review.



Get started with AWS AI services for content moderation

The volume, complexity, and speed of UGC will continue to increase, and technical complexity plus customer expectations and regulations worldwide will continue to put pressure on organizations to keep communities safe and inclusive. Eliminating harmful and unwanted content leads to increased business growth and opportunities.

ML/AI can help organizations achieve their objectives while reducing costs and human moderation workloads. AWS has AI services that can be self-managed or customized to help you create safe online communications and protect your customers and brand, all while reducing moderation costs.

[🚀 Get started for free ›](#)

Getting started is easy. Your team can try any of these services for free, implement them with no ML experience, or work with one of our AWS Partners with proven experience in delivering operational efficiency and cost savings for customers worldwide. AWS makes content moderation easy, from free trial to proof-of-concept (POC) and all the way through production.

[🚀 Connect with an AWS expert now ›](#)

AWS experts, solution architects and partners can help you with a POC and assistance from AWS sales, technical teams, or AWS Partners is also a click away. The AWS free tier makes starting free and easy, enabling your best innovation with speed.

[🚀 Learn more about AWS AI services and solutions for content moderation ›](#)