

Content Moderation services on AWS



Text



Amazon Comprehend
Trust and Safety



Amazon Bedrock
Guardrails



Amazon Bedrock LLM
with Classifying prompt



Amazon SageMaker



Image



Amazon Bedrock
Data Automation



Amazon Rekognition
Content Moderation



Amazon Bedrock
Guardrails



Video



Amazon Bedrock
Data Automation



Amazon Rekognition
Content Moderation

Content Moderation services on AWS – Text & Documents

Options Features	Amazon Comprehend Trust and Safety	Amazon Bedrock Guardrails	Amazon Bedrock LLM with Classifying prompt	Custom Random Forest Classifier model
Modality	Text	Text / Image	Text	Text
Evaluated input	Text segments	LLM Input / Output Standalone Text / Image	Prompt with LLM supported modalities	Embeddings
Input constraints	Up to 10 segments x 1KB	Up to 25K chars/second per policy (Standalone text)	Per selected LLM constraints	
Moderated categories	8 Moderated categories	5 harmful categories + Custom denied topics	Custom moderated categories	Toxic / Non-toxic
Toxicity dataset	Built-in dataset	Built-in dataset / Custom phrases/file	LLM-knowledge driven	Custom dataset
Confidence score	Y	Y	N	Y
API	DetectToxicContent	LLM Prompt/Content: InvokeModel / Converse Standalone Text/Image: ApplyGuardrail	InvokeModel / Converse	
Batch API	NA	NA	CreateModelInvocationJob	
PII/PCI handling	Detect / Redact via dedicated APIs	Detect by type, regex / Block / Mask	As instructed	
Cost	Per characters	Per characters	Per in+out tokens	Solution-based
Unit price (N. Virginia)	\$0.0001 / 1B chars	\$0.15 / 1M chars		
Language	English	English, French, Spanish		
Regions	N. Virginia, Oregon, Sydney, Ireland	Supported regions	Supported regions	
Additional features	Prompt safety (explicit or implicit malicious intent: discriminatory, illegal, unsolicited content)	Profanity filter, Prompt attacks, Denied topics		

Content Moderation services on AWS – Images

Options Features	Amazon Bedrock Guardrails	Amazon Bedrock Data Automation	Amazon Rekognition Content Moderation
Modality	Image / Text	Image	Image
Evaluated input	LLM Input / Output Standalone Image / Text	S3 bucket	Base64-encoded blob / S3 file
Input constraints	Up to 25K chars/second per policy (Standalone text)	JPEG, PNG S3 file: up to 15 MB	JPEG, PNG Blob: up to 5 MB S3 file: up to 15 MB
Moderated categories	5 harmful categories + Custom denied topics	7 moderation categories	3-level hierarchical categories
Toxicity dataset	Built-in dataset / Custom phrases/file	Built-in dataset	Built-in / Custom via adapters
Confidence score	Y	Y	Y
API	LLM: InvokeModel / Converse Standalone: ApplyGuardrail	InvokeDataAutomationAsync GetDataAutomationStatus	DetectModerationLabels
Batch API	NA	InvokeDataAutomationAsync GetDataAutomationStatus	StartMediaAnalysisJob GetMediaAnalysisJob
PII/PCI handling	Detect by type, regex / Block / Mask		
Cost	Per characters	Per image/video min	Per image
Unit price (N. Virginia)	TBC	\$0.003/image	\$0.001 / first 1M images \$0.0008 / next 4M images
Language	English, French, Spanish		
Regions	Supported regions	N. Virginia, Oregon	Supported regions
Cross-region inference		Y	N
Additional features	Profanity filter, Prompt attacks, Denied topics	Image summary, Logo detection, IAB taxonomy, Image text detection	Face & identity detection, Label & text detection

Content Moderation services on AWS – Video

Options Features	Amazon Bedrock Data Automation	Amazon Rekognition Content Moderation
Modality	Video	Video
Evaluated input	S3 bucket	S3 file
Input constraints	MOV, MKV up to 2GB	MPEG-4, MOV up to 10GB
Moderated categories	7 moderation categories	3-level hierarchical categories
Toxicity dataset	Built-in dataset	Built-in / Custom via adapters
Confidence score	Y	Y
API	InvokeDataAutomationAsync GetDataAutomationStatus	StartContentModeration
Batch API	InvokeDataAutomationAsync GetDataAutomationStatus	NA
PII/PCI handling		
Cost	Per image/video min	Per video min
Unit price (N. Virginia)	\$0.050/minute	\$0.10/min
Language		
Regions	N. Virginia, Oregon	Supported regions
Additional features	Logo detection, IAB taxonomy, full video/ video chapter/ img summarization, text recognition, video transcription	Face & identity detection, Label & text detection