# AWS AI Services

Rekognition & Lex – Simplified Deep Learning Solutions

Rekognition is and AWS offering which help to embed visual analysis to various applications. It is capable to searching, verifying and organizing tens of thousands of images. It is also capable to analyzing motion-based content. It is developed on highly scalable deep learning technology which allows detection of objects, faces, scenes and read textual content, inappropriate content identification and it can also do face comparison.
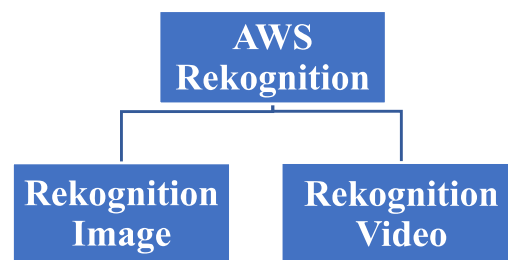
Rekognition Image detects any objects and scenes in images. Rekognition video detects movement of objects in a frame. It is capable to detect human movement even when the face is not visible in frame. This makes up for a use case to get a notification whenever a delivery person comes near your entry gate for parcel delivery. Some of the widely implemented use cases of Rekognition Image are face recognition, sentimental analysis and those of Rekognition Video are search indexes and explicit content filtering in online media.

The best part about this service is that one can use deep learning without having the expertise of building, maintaining or upgrading pipelines. For better accuracy of any computer vision task, huge amount of labelled ground truth data should be used for training along with Graphics Processing Units (GPU) to allow high computations. But all this is taken care of when using Rekognition. It comes pre-trained for recognition related tasks so the deep earning pipeline is fully managed. This helps to keep our focus on design and development of the core application. The following figure shows the functions supported by Rekognition service.

1. Object and Scene Detection: Thousands of object models are pre-trained for recognition in Rekognition. So, for labelling of objects in any image, this API is used.
2. Custom Labelling: It is not for detecting faces, text but for finding custom objects and scenes in any image.
3. Content Moderation: It is an important API which allows detection of explicit content having violence, weapons, any sort of disturbing content like alcohol, drugs etc. It is mostly used in child apps, dating sites, photo sharing apps etc.
4. Facial Analysis: It allows to detect faces from images. It is also able to detect attributes like sunglasses, gender etc.

Object and Scene Detection	**People Pathing**
Custom Labelling	**Black frames Detection**
Content Moderation	**End credits Detection**
Facial Analysis	**Shot Detection**
Face Comparison	**Color Bars Detection**
Text Detection	**Facial Recognition**
PPE Detection	**Celebrity Recognition**

5. Face Comparison: It allows to compare two faces to know about their similarity levels. It is used to match real person with photos.

6. Facial Recognition: It allows to identify any person by matching the captured frame from a collection or group of face frames.

7. Celebrity Recognition: It is used to identify certain famous personalities in the field of sports, politics, entertainment, business etc.

8. Text Detection: It helps to recognize the text in any image/video like addresses, captions, subtitles etc. Visual search on social media and vehicle plate number identification by traffic cameras are its significant use cases.

9. PPE Detection: This is an interesting offering which detects different types of face, hand and head covers in the covid scenario.

10. People Pathing: Rekognition allows you to keep track of people's paths which allows generation of customer insights like their movement pattern in aa shopping mall, grocery store.

11. Black frames Detection: Usually any video having ads will have black frames/empty frames to be used as cues for ads to begin. Using Rekognition, this black frame insertion process is automated.

12. End credits Detection: Rekognition allows to precisely identify when the ending credits section begin in any movie so that prompts like "next movie" can be timely shown.

13. Shot Detection: A set of related images taken continuously is called a shot. Rekognition allows detection of starting, ending and duration of every shot. This is useful to setup thumbnails during video transition and identify spots for ads to begin.

14. Color Bars Detection: Rekognition helps to detect sections in a video which comprise of SMPTE bars which ensure color is accurately calibrated. It is useful in detection of issues namely broadcast signal loss.

	Google Vision	Microsoft Computer Vision	IBM Watson	AWS Rekognition
Face Detection	Yes	Yes	Yes	Yes
Face Recognition	No	Yes	No	Yes
Face Landmarks	Yes	Yes	No	Yes
Feature Detection	Yes	Yes	No	Yes
Similar Faces	No	Yes	No	Yes
Emotion	Yes	Yes	No	Yes
Label Detection	Yes	Yes	Yes	Yes
Landmarks	Yes	Yes	Yes	No
Celebrities	No	Yes	No	Yes
Logo Detection	Yes	No	No	Yes
NSFW	Yes	Yes	Yes	Yes
Image Analysis	Yes	Yes	Yes	Yes
Video Analysis	Yes	Yes	No	No
Custom Model Creation	No	Yes	Yes	Yes

Fig. Comparison among top computer vision APIs by public cloud providers.

Amazon Lex

In today's world, the chatbot market is growing expansively and almost every kind of business is benefitting from it. Majorly the chatbot can be of three types, namely- rule based, AI bots and hybrid bots. The rule-based/linguistic chatbots offer fine-tuned control and are highly flexible. They make use of if-else logical conditions to direct conversation flow using a linguistic model which is why their interaction capabilities are very structured. One can easily come across these chatbots on e-commerce platforms and social networking sites. The AI bots, though higher on complexity, they offer a more real-life like conversation. Over time they learn and become contextually better. Due to machine learning capabilities it learns from earlier conversations. Then comes the Hybrid model which takes up the best of both- rule based and AI. It makes use of ML integrations which go beyond linguistic rules.

Amazon Lex is an offering for making conversational bots which are capable of interacting through voice and text. It is backed by ASR (Automatic Speech Recognition). Using lex these chatbots can be published to various chat services and mobile devices. The best part about using Lex id that machine learning expertise is not required for developers. The language model is automatically built by prompts that are inputed. There are several use cases like information bots, bots to control devices, order placing or travel bots, self-service bots etc. There are no bandwidth constraints and Lex auto-scales as per your needs. It uses deep learning to get smarter with time.

Lex has iOS and Android SDKs for mobile development and one does not need to certify the bot before deployment. The maximum speech input time is 15 seconds currently for slot filling and the languages supported are US Spanish, Canadian French, British English, Australian English, German, US English, Latin American Spanish and French. Lex V2 has an additional streaming API in which the bot will listen continuously and respond proactively. Messages like "Take your time to respond" can be further added to make the conversation human-like.

As per Gartner report of 2019 (source: https://www.gartner.com/en/documents/3953723/market-guide-for-conversational-platforms) the following are the top 16 market vendors for conversational chatbots- Avaamo, Amazon Web Services, Artificial Solutions, Eudata Google, IBM, IPsoft, Kore.ai, Microsoft, OneReach.ai, Openstream, Oracle, Rasa, Rulai, SmartBotHub, SoundHound. Other options like Gensim, Textblob, PyNLPL, CoreNLP, spaCy Python NLP libraries are also widely used to build AI chatbots.

	Amazon Lex	IBM Watson	Microsoft Bot
Ease of Use	Yes	Yes	Yes
Integrations	SMS, Slack, FB Messenger, Kik, Twilio	Voice Agent, Slack, FB Messenger, Wordpress, Custom APIs	FB Messenger, Slack, Skype, Kik, Telegram, Twilio
Web and Mobile Integrations	Basic chat widget UI for testing on website	Basic chat UI available for websites	Open source chat widget available
Languages	Spanish, Canadian French, British English, AU English, German, French, US English etc.	English, Spanish, Japanese, Chinese, Italian etc.	English, French, German, Spanish etc.
Pricing	$0.004/ voice request $0.00075/ text request	$0.14 per user per month	$0.0005/message request

Fig. Comparison of major chatbot building frameworks

Use Cases

➤ Automated Covid assessment bot

The client requirement was having a Covid assessment tool which consists of a genie that interacts as a chatbot for voice and text-based interactions. It leverages multiple features of AI such a voice, Natural language processing and decision making to provide a solution that can scales up to serve multiple people. Its developed using Amazon Lex interacts with the user and suggested whether he or she is prone to COVID-19 based on a set of related questions, the bot is conversational and responds to user's voice and text replies in real time. It is accessible on the following link https://master.dcjrsebf77vbs.amplifyapp.com/Chat
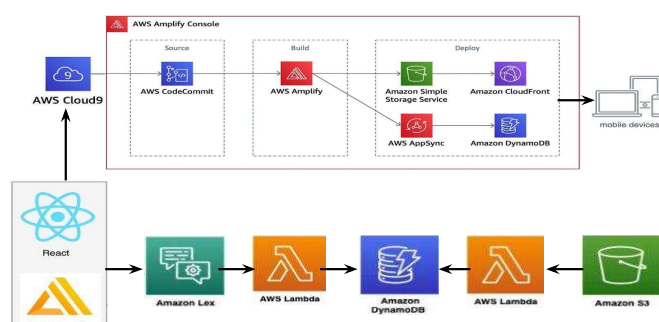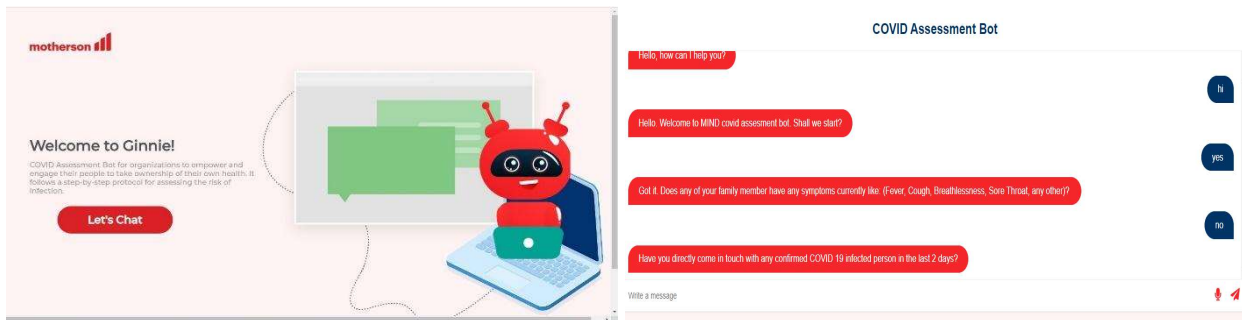
Fig. Deployed COVID Chatbot

➤ 3D, augmented reality (AR) and virtual reality (VR) application

The customer required a virtual environment consisting of an interactive chatbot for voice and text-based interactions. It includes the real time conversations with the bot and the user can also see the captioning of the responses which improves the user experience.

The scene consists of a wall television which plays a video and adds to the aesthetics of the environment. The host of the scene in Sumerian is in sync with Amazon Lex – which is used for creating the bot here, it lets us build the conversational interface and build chatbots without any heavy lifting. Amazon Polly then turns the text response by Lex into speech. Lambda function is also written in python to act as the backend of the whole task by initializing and validating the user input.
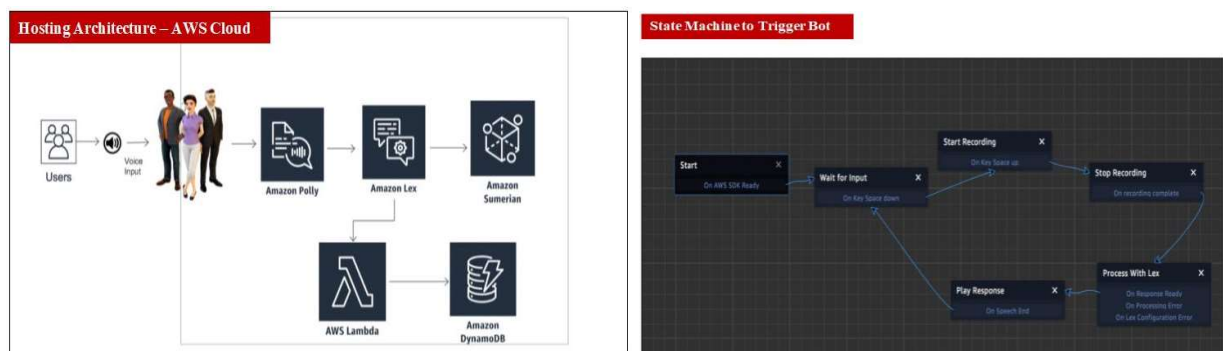
Fig. A close view of the VR scene

➤ **Enterprise level Attendance tracking system**

The customer required an attendance tracking system for its employees. For this purpose, we used the face recognition capabilities of AWS Rekognition along with services DynamoDB and Lambda. A collection of image library is created in Rekognition which comprises of training images for every employee, corresponding these set of images the video frame captured is matched for attendance marking of that employee. S3 is used to store the collection. The workflow comprises of two sections- Indexing and Analysis.

Indexing comprises of populating the collection in Rekognition using IndexFacesAPI and Analysis comprises of queries that will be run on this collection using SearchFacesByImage API. DynamoDB is used to store the key-value pair like "emp_id-present" for later reference and UI display. Lambda function is written in python to act as the backend of the service interactions.
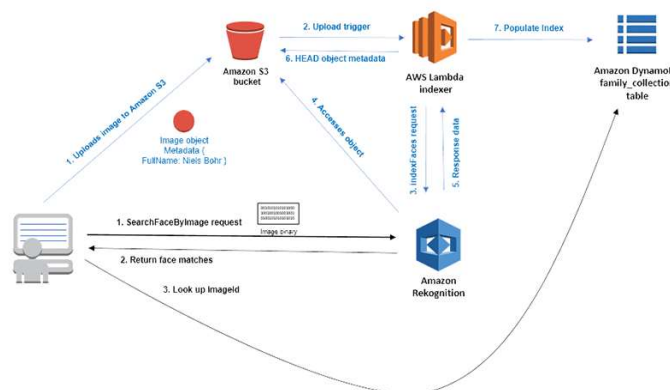

Fig. Architecture diagram