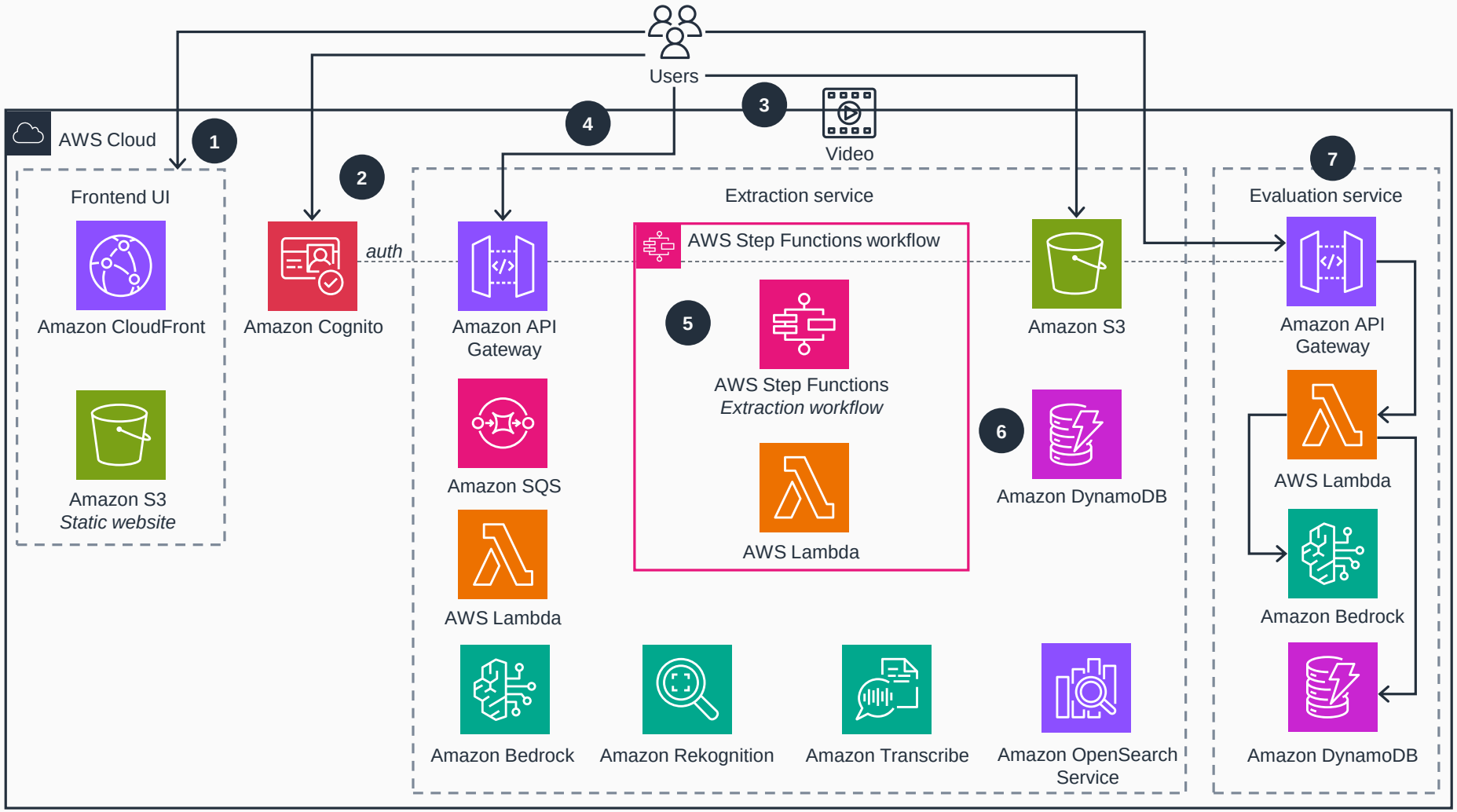
Guidance for Media Extraction and Dynamic Content Policy Framework on AWS

Extraction of generic metadata

This architecture diagram shows how to use generative AI to extract generic metadata from videos and demonstrates a dynamic policy evaluation analysis.



1. Media analysts access front end static website through **Amazon CloudFront** distribution. The static content hosted on **Amazon Simple Storage Service (Amazon S3)**.
2. Users log in to the frontend web application, authenticated by an **Amazon Cognito** user pool.
3. Users upload video(s) to **Amazon S3** directly from the browser using multi-part, pre-signed **Amazon S3** URLs managed by the UI application.
4. The frontend UI interacts with the extract service (microservice) through a RESTful interface provided by **Amazon API Gateway**. This interface offers Create, Read, Update, Delete (CRUD) features for video task extraction and management. The extraction service can be deployed and used independently of the other components.
5. An **AWS Step Functions** state machine oversees the analysis process. It transcribes audio using **Amazon Transcribe**, samples image frames from video using moviepy, uses multimodal models on **Amazon Bedrock** to analyze images, and uses **Amazon Rekognition** for additional insights. It also generates text and multimodal embeddings on the frame level. Users can customize the logic in this Guidance to integrate their preferred generative AI models.
6. **Amazon DynamoDB** stores media processing task metadata and extracted video information in text format. An **Amazon OpenSearch Service** cluster stores vector embeddings and facilitates search and discovery needs.
7. Using the solution UI, the user selects and customizes existing template prompts, then initiates the policy evaluation utilizing **Amazon Bedrock** large language models (LLMs) based on the extracted video metadata.

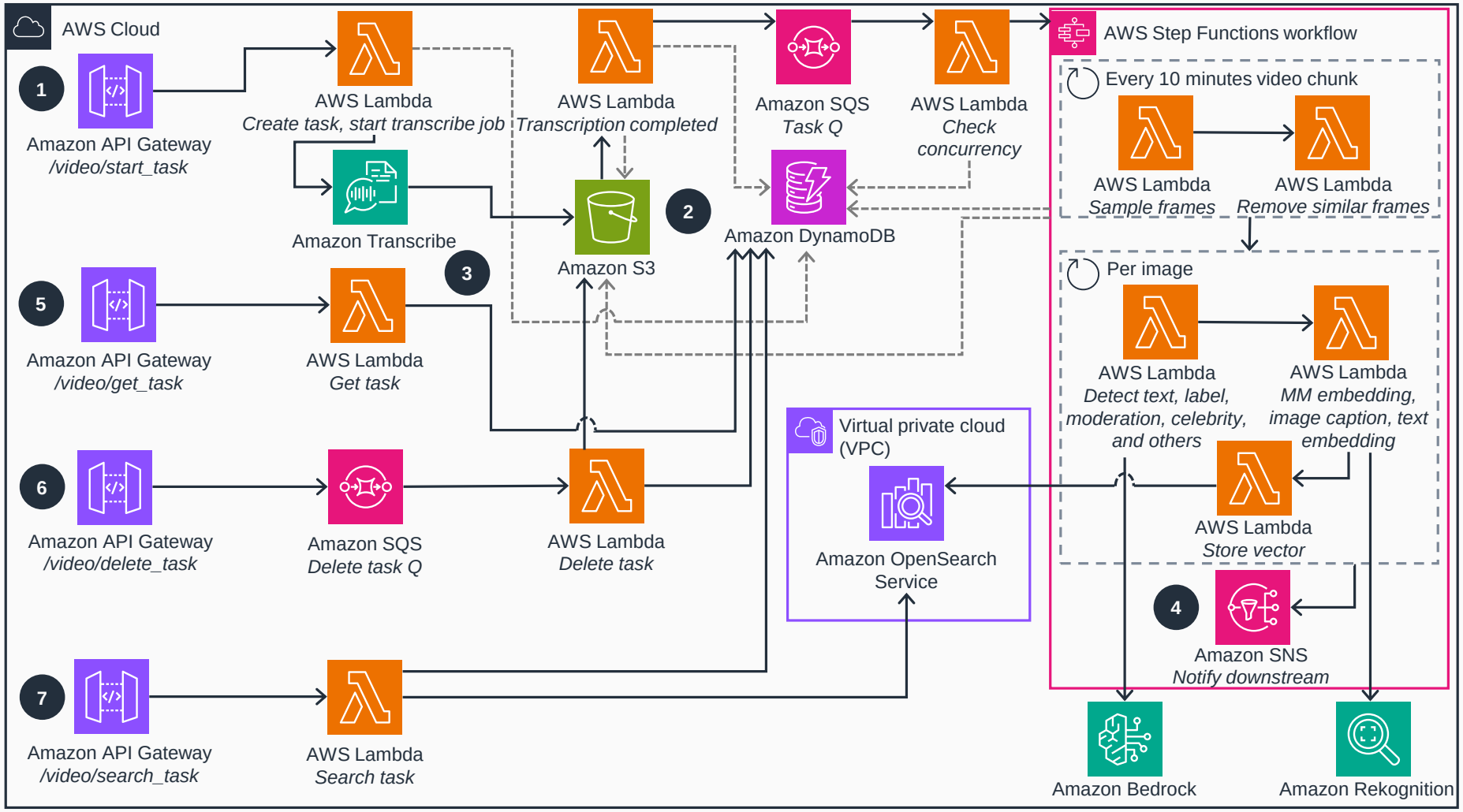
Reviewed for technical accuracy February 11, 2025

AWS Reference Architecture

Guidance for Media Extraction and Dynamic Content Policy Framework on AWS

Restful APIs of the extraction service

This architecture diagram illustrates the key RESTful APIs of the extraction service, served through Amazon API Gateway. The UI uses APIs to retrieve data, allowing users to integrate the extraction service into existing workflows.



1. The */start_task* endpoint serves as the core of the extraction service, managing the video metadata extraction process and maintaining the results.
2. **DynamoDB** stores the extracted metadata. The raw results from generative AI models are saved as JSON or text files in **Amazon S3**. **Amazon OpenSearch Service** indexes store frame-level embeddings to serve search.
3. The process includes invoking **Amazon Transcribe** to generate audio transcriptions, sample image frames from the video at a specified interval, and remove similar frames by generating multimodal embeddings and applying similarity comparison. For each image frame, the service applies AI or generative AI features to extract metadata. Additionally, the service generates text and multimodal embeddings for each frame to enable vector search capabilities.
4. **Amazon Simple Notification Service (Amazon SNS)** notifies downstream workflows of task completion.
5. The */get_task* endpoint retrieves video task information using a unique task ID. The data is fetched from the **DynamoDB** tables.
6. The */delete_task* endpoint deletes video tasks using a unique task ID. It will delete all the task-related states from **DynamoDB** tables, **Amazon S3**, and **Amazon OpenSearch** indexes.
7. The */search_task* endpoint searches for tasks matching the provided criteria. It supports keyword searches against the **DynamoDB** task name and description, as well as semantic and multimodal embedding searches using the **Amazon OpenSearch** vector index.

Reviewed for technical accuracy February 11, 2025

AWS Reference Architecture