SCTE | STANDARDS

Data Standards Subcommittee

SCTE STANDARD

SCTE 276 2022

Video Metadata Extraction in Digital Advertising

NOTICE

The Society of Cable Telecommunications Engineers (SCTE) Standards and Operational Practices (hereafter called "documents") are intended to serve the public interest by providing specifications, test methods and procedures that promote uniformity of product, interoperability, interchangeability, best practices, and the long term reliability of broadband communications facilities. These documents shall not in any way preclude any member or non-member of SCTE from manufacturing or selling products not conforming to such documents, nor shall the existence of such standards preclude their voluntary use by those other than SCTE members.

SCTE assumes no obligations or liability whatsoever to any party who may adopt the documents. Such adopting party assumes all risks associated with adoption of these documents and accepts full responsibility for any damage and/or claims arising from the adoption of such documents.

NOTE: The user's attention is called to the possibility that compliance with this document may require the use of an invention covered by patent rights. By publication of this document, no position is taken with respect to the validity of any such claim(s) or of any patent rights in connection therewith. If a patent holder has filed a statement of willingness to grant a license under these rights on reasonable and nondiscriminatory terms and conditions to applicants desiring to obtain such a license, then details may be obtained from the standards developer. SCTE shall not be responsible for identifying patents for which a license may be required or for conducting inquiries into the legal validity or scope of those patents that are brought to its attention.

Patent holders who believe that they hold patents which are essential to the implementation of this document have been requested to provide information about those patents and any related licensing terms and conditions. Any such declarations made before or after publication of this document are available on the SCTE web site at https://scte.org.

140 Philips Road
Exton, PA 19341

Document Types and Tags

Document Type: Operational Practice

Document Tags:

☐ Test or Measurement	☐ Checklist	☐ Facility
☐ Architecture or Framework	☐ Metric	☐ Access Network
☒ Procedure, Process or Method	☒ Cloud	☐ Customer Premises

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. Executive Summary

Video is a ubiquitous medium. From TV and movies to social media and mobile platforms, its applications are numerous in the entertainment sector. The ability to 'describe' what's happening in a video, without human intervention is the eventual goal of an AI/ML engine. The usual definition of Metadata is 'data about data'. However, in the context of machine learning, Metadata are the 'content descriptors' of video/audio/textual data elements extracted from a video. Video metadata extraction is the process of auto-identifying video content.

An emerging trend is the application of AI technology for TV advertising. In this Best Practices Guide, we discuss the unique challenges in applying machine learning to carrier-class video advertising. To illustrate the point, the discussion is focused on a specific use case that is common to all ad supported TV services.

The selected use case is Ad Ingest Quality Control (QC). TV commercials are subjected to various rules and regulations. For example, ads containing specific content (e.g. Alcohol, firearms) are barred from airing during certain TV programs. Identifying these categories might pose a challenge to a machine learning tool, as off-the-shelf products are more oriented towards facial recognition. That is to be expected perhaps, as the video ML products were primarily intended for surveillance and sports applications. However, by judiciously combining metadata from multiple data streams, ML based analysis can be enhanced.

1.2. Scope

This best practices guide consists of two parts. The solution description section contains recommendations based on AI/ML WG member experiences and the lessons learned. The requirements section defines features a machine learning tool would need to perform for specified tasks.

1.3. Benefits

Machine learning based video analysis is a burgeoning field. As the technology matures, its applications in broadband industry will be far and wide. Network operators and service providers will find the guidelines useful in solution development. Vendor partners who are developing carrier-class machine learning solutions might embody these in product specifications.

1.4. Intended Audience

The intended audience are the machine learning practitioners in cable-telecom space.

1.5. Areas for Further Investigation or to be Added in Future Versions

While digital advertising is the focus of the current document, the methodologies are applicable to general video analysis as well. As the technology matures, guidelines to perform semantically rich contextual analysis would be fruitful.

Emerging paradigms, such as Interpretable and Explainable Artificial Intelligence (XAI), challenge the 'black box' model of deep learning. The extracted video metadata would need to conform with the results of such analyses.

2. Normative References

The following documents contain provisions which, through reference in this text, constitute provisions of this document. The editions indicated were valid at the time of subcommittee approval. All documents are subject to revision and, while parties to any agreement based on this document are encouraged to investigate the possibility of applying the most recent editions of the documents listed below, they are reminded that newer editions of those documents might not be compatible with the referenced version.

2.1. SCTE References

No normative references are applicable.

2.2. Standards from Other Organizations

No normative references are applicable.

2.3. Other Published Materials

No normative references are applicable.

3. Informative References

The following documents might provide valuable information to the reader but are not required when complying with this document.

3.1. SCTE References

No informative references are applicable.

3.2. Standards from Other Organizations

No informative references are applicable.

3.3. Other Published Materials

[1] FCC Guidelines for Ads - https://www.fcc.gov/consumers/guides/complaints-about-broadcast-advertising

[2] FTC Guidelines for Ads - https://www.ftc.gov/tips-advice/business-center/guidance/ftcs-endorsement-guides-what-people-are-asking

[3] FEC Guidelines for Ads - https://www.fec.gov/help-candidates-and-committees/making-disbursements/advertising/

[4] FDA Guidelines for Ads - https://www.fda.gov/media/82590/download

[5] ESPN Advertising Guidelines - http://www.espn.com/adspecs/guidelines/en/ESPN_AdStandardsGuidelines.pdf

[6] MIT AI Lab research - http://moments.csail.mit.edu/explore.html

4. Compliance Notation

shall	This word or the adjective "***required***" means that the item is an absolute requirement of this document.
shall not	This phrase means that the item is an absolute prohibition of this document.
forbidden	This word means the value specified ***shall*** never be used.
should	This word or the adjective "*recommended*" means that there *may* exist valid reasons in particular circumstances to ignore this item, but the full implications *should* be understood and the case carefully weighed before choosing a different course.
should not	This phrase means that there *may* exist valid reasons in particular circumstances when the listed behavior is acceptable or even useful, but the full implications *should* be understood and the case carefully weighed before implementing any behavior described with this label.
may	This word or the adjective "*optional*" indicate a course of action permissible within the limits of the document.
deprecated	Use is permissible for legacy purposes only. Deprecated features *may* be removed from future versions of this document. Implementations *should* avoid use of deprecated features.

5. Abbreviations

AI/ML	artificial intelligence/machine learning
ANN	artificial neural network
CNN	convolutional neural network
DL	deep learning
FED-ML	federated learning
FN	false negatives
FP	false positives
JSON	JavaScript Object Notation
KNN	k-nearest neighbor
LSTM	long short term memory
MVPD	multi-channel video programming distributors
OTT	over the top
QC	quality control
R-CNN	region based convolutional neural network
RNN	recurrent neural network
SVM	support vector machines
TFLOP	trillion floating-point operations per second
V-MVPD	virtual MVPD
XAI	explainable AI

6. Broadcast Advertising – Constraints

Multi-channel video programming distributors (MVPD) are highly regulated in the US, for example. The term covers not only traditional cable companies, but any entity that provides TV service to consumers via fiber, coax, satellite, DSL or wireless. With the advent of internet-based TV service (also known as

over the top (OTT)), the moniker is modified as V-MVPD (virtual MVPD). Note that in all cases, the content distributors could be responsible for the displayed video content, including advertisements [1] . This places the onus on the content distributor (also known as service provider/network operator), to prevent the "non-compliant" content from reaching the TV audience.

In the context of the present discussion, there is a distinction between TV content and ads. While movies/episodes etc. are originated from mainstream studios (and thus properly vetted), TV ads could originate from a multitude of sources. Therefore it is necessary to identify any non-compliant ads at the Ad Ingest Quality Control (QC) prior to airing. Today, this is done manually by trained individuals. They examine tens of thousands of ads a month and quarantine the failed ones. The challenge is to automate that process with an AI/ML engine embedded within the workflow.

First, we examine the basis for non-compliancy of ads. While reference is made to the US regulatory framework, similar laws apply in other countries. When a TV commercial is deemed non-compliant, the restriction usually stems from one of the three categories below.

a) Regulatory Compliance

The regulatory constraints are primarily stipulated by FCC [1] , but could also be under the purview of FTC, FEC and FDA [2] [3] and [4] Listed below are some examples of regulatory requirements overseen by federal agencies. See the references cited above for full requirements.

Examples:

- "Broadcasters are responsible for selecting the broadcast material that airs,…including advertisements." [1]
- Ads related to alcohol, tobacco, firearms, gambling, etc. must meet federal guidelines.
- A political ad is required to display a statement from the sponsor for at least 4 seconds.
- An ad might be deemed deceptive for misleading/missing information (truth-in-advertising).
- Ads promoting certain lotteries, cigarettes or smokeless tobacco products are not allowed.
- Ads must comply with loudness mitigation requirements of the CALM Act.

b) Contractual Compliance

Contractual constraints are imposed by content providers such as ESPN. An example would be the restriction on alcohol ads during ESPN Little League World Series program. For a complete list of applicable restrictions, see reference [5]

c) Business/Operational Compliance

These are generally operational guidelines and best practices established by the broadcasters. Being sensitive to audience needs as well as delivering quality content could enhance a company's credibility. One example is "frequency capping" or limiting the display of the same ad multiple times.

7. ML Challenges in Carrier-Class Video Analysis

Identifying the above categories programmatically might pose a challenge to ML tools, as off-the-shelf products are more oriented towards facial recognition. A familiar ML application is creating a "bounding box" around a face and tracking it through a video clip (Figure 1). Such applications are useful in sports and surveillance; however they are not directly applicable to the MVPD market. The latter requires comprehensive ML analyses of multiple streams (video, audio and textual metadata).

Figure 1- Bounding Box Example

(Photo Credit: Pexels.Com)

7.1. ML Based Video Metadata Extraction – Tool Capabilities

Use of machine learning (ML) for image and video analysis often include face recognition, personalization and recommendations. The list below shows the general capabilities of ML tools in the market.

a) Video/Image

- Face recognition
- Object detection
- Activity identification
- Emotions (smiling/frowning)
- Celebrity identification

b) Audio

- Specific phrases
- Sentiment (positive or negative)

c) Text/OCR

- Transcription of the audio
- Generation of text from product labels

In common usage, machine learning video products do a multi-pass analysis with each pass identifying specific characteristics, such as faces, common objects, celebrities etc. The results are presented as content descriptor metadata (labels). An accompanying "confidence level" indicates the accuracy of prediction. Off-the-shelf ML tools might not meet the needs right out of the box, as the video content/ad detection is still a nascent technology. Customizing such products for carrier-class video applications requires a certain amount of post-processing. Otherwise, the results could be tainted with false positives or the tool might fail to identify content adequately (false negatives).

8. Deep Learning Algorithms for Video Classification

Object detection and recognition are classification problems in machine learning. Compared to image analysis, video is inherently complex. While image analysis has only spatial dependence, video analysis adds the temporal component. The statistical algorithms used mainly are support vector machines (SVM),

decision trees and k-nearest neighbor (KNN). The current trend however is for neural network-based algorithms. This is mainly due to two drivers: prevalence of large amounts of data for training and the availability of high speed GPUs for parallel computing.

The generic artificial neural networks (ANN) models suffer from accuracy and training time issues and are not able to meet carrier-class video classification requirements. Hence more sophisticated deep neural networks have become the standard. Of these, the convolutional neural network (CNN) is the workhorse and is a powerful image classification tool. While a video is composed of a succession of images, the time dimension makes the analysis more complex.

Recurrent neural networks (RNN) algorithms have been successfully applied to this task. There have been several variants of RNN over the years. For time series analysis, RNN-based deep learning models are the standard technique due to the ability to store events happened in the past. While RNN is capable of analyzing sequences of data, training a machine learning model based on RNN is a formidable task. That's because it is susceptible to "vanishing/exploding gradient" problems during the training phase, when a back-propagation technique is applied. The root cause is the exceedingly small derivative of the "loss function" (or error) during back propagation. To avoid extreme values of the gradients, it is necessary to disregard certain intermediate steps. A solution for this problem is a Long Short-Term Memory (LSTM) algorithm, which is a modified version of RNN. LSTM adds the capability to remember longer time steps without issues, via the use of multiple gates. The flip side is that LSTM is complex to compute.

AI/ML-based video analysis is a burgeoning field with new and improved algorithms. New algorithms are routinely being developed (Fast R-CNN, Faster R-CNN, etc.). These are mainly for improving the speed of analysis as updating millions of parameters (weights and biases) associated with hidden states takes a lot of time.

8.1. Machine Learning Paradigms

Several machine learning paradigms are discussed below.

Transfer Learning – Transfer learning is the reuse of a pre-trained model on a similar, but new problem. The ML engine is first trained with a publicly available dataset and then fine-tuned to suit the specific application. This technique speeds up the weight initialization process and reduces training time of the neural network model.

Federated Learning – The core concept of Federated learning is decentralized learning, meaning that the data stays within each device/domain. This new paradigm is especially suited for training wireless devices as it seems to address the privacy concerns. The algorithm is trained incrementally and locally on the device. The updates are sent periodically to a central server.

Understanding the intricacies of Federated Learning (Fed-ML) would be crucial for service providers. Some of the pertinent questions one might ask are:

- What are the privacy/security loopholes?
- What is the defense against backdoor attacks?
- Are certain models more susceptible for data breaches?
- What are the risks of data reconstruction?
- What service provider obligations exist in a multi-domain Fed-ML model?
- What are the patent encumbrances for operators that deploy Fed-ML?

Explainable AI (XAI) – Deep learning models function more or less as black boxes. A neural network can easily classify a photo of an animal as a cat but might be reticent about why it made that decision. A recent development is explainable AI (XAI), also known as interpretable-AI (with different nuances). The intent is to help open up the black box model. For example, XAI could provide the justification as to why an ML engine has declined a loan application or why a specific product was recommended for a specific customer.

In the case of video metadata usage, the applications span a wide range. The extracted video metadata *should* align with the aforementioned machine learning paradigms.

8.2. Training a Neural Network

To train a neural network, a good selection of examples and counter-examples is needed, else the machine learning model would be susceptible to overfitting. That is, the model will fit the existing data well, but is likely to fail when it encounters a new instance of the target data. While this is not an issue with common objects (e.g. cars) due to the abundance of examples, it is a challenge for objects with ambiguous signatures (such as fireworks or alcohol). Distinguishing fireworks from similar signatures (bright lights in a dark background), is not an easy task. Similarly, an image classifier might find it hard to differentiate beer from a similarly colored liquid in a bottle (e.g. olive oil).

The need for proper counter-examples becomes more acute as we move from image analysis to video activity identification. This is discussed in detail in the Error Analysis section.

8.3. Test Planning

During the test planning stage, an assessment needs to be made on the number of classes and samples. The similarities and variations/imbalances in classes can affect the engine performance. Generally, the dataset is split in 70:20:10 ratio among the training, test and validation data. It is also possible to forego a separate data set for validation (only 80:20) via k-fold cross-validation with the training dataset. The data set also needs to be carefully assessed to ensure that it is balanced. Unbalanced data can skew the model predictions. Oversampling (duplication) and under-sampling (random deletions) are standard techniques for correcting imbalanced data.

Fine tuning the model and hyper-parameters yields improved performance of the ML engine. A model-parameter example is neural network weights optimization. Hyper-parameters would include the number of layers, learning rate, number of neurons per hidden layer, etc.

9. Machine Learning Tool Performance

While ML engine performances continue to improve, current detection speeds are slower than real time. For certain tasks, the ML engine could take excessive time for the video analysis (e.g. many minutes for a 30-second ad). One reason could be that the ML engines operate in multi-pass mode. This is necessary because at Ad-Ingest Quality Control, the ML engine works as a gate-keeper. On the other hand, if the intent is to find a single signature (e.g. either guns or alcohol), a single pass would be sufficient.

To improve the speed, one approach would be to reduce the number of categories selected for detection. The ML engine has to perform a classification task per each category. Some categories might not be relevant to the task at hand, although each task consumes time for detection. Another option would be to use faster GPU processors.

9.1. Hardware Considerations

Machine learning engines can be appliances or reside in the cloud. The cloud-based implementation is preferred if the data also resides in the cloud. The appliances would be GPU-based (as opposed to CPU), due to the large number of cores which facilitate parallel computing.

When it comes to benchmarking, NVidia GTX (or similar) products are general purpose engines. For specialized applications NVidia DGX (with thousand TFLOPs of computing speed) might be a choice. Benchmarking a cloud product is trickier due to multiple factors that can affect the performance. Analysis based on statistical results is recommended.

10. Resurgence of Contextual Advertising

In the last two decades, cookie-based user tracking was the primary mode for targeted advertising in digital media. However, the latest privacy regulations (GDPR, CCPA) are making such data collection practices unacceptable. Therefore, advertisers are eager to find alternative means to promote their brands. Contextual advertising is a suitable option as it protects consumer data. However, that necessitates powerful AI/ML capabilities to describe a scene in a video image. Generating descriptive metadata is a challenge. For example, instead of generic labels such as "person/human," the ML tool needs to identify whether a person is young/old, male/female, mood, activity, etc.

10.1. Video Analysis for Contextual Advertising

Activity identification is a burgeoning field of research [6] Though steadily improving, it is a challenge for the ML tools. Content descriptors (labels) need to be sufficiently descriptive for effective contextual analysis. For the MVPD space, "activity identification" would open up new applications such as identifying a car chase from a video (as opposed to cars in a still image) would offer new ad opportunities. Table 1 below depicts sample activities that are relevant to contextual advertising.

Table 1 - Activity Identification for TV Advertising (Examples)

Dominant Activity	Possible Ad Usage
Cooking	Utensils, cooking classes, kitchen appliances
Car chase	Car ads/repairs, auto insurance
Shopping	Retail store ads
Eating	Food ads, restaurant
Dancing	Clothing, personal care, alcohol ads
Drinking	Alcohol ads
Social gathering	Multiple products
Kids playing	Toys, food and drink ads, medicines, clothing
Sports activities	Sports-related products
Anxiety, Arguing	Pain medications, lawyer ads

11. Error Analysis of the ML Classifier

Measuring the accuracy of an ML classifier is not a straightforward task. The usual definition of accuracy (the number of correct results divided by the total results) might not yield a useful measure. This is due to imbalances in the dataset. To understand this better we need to look at the types of errors to which classifiers are susceptible.

- False Positive – Incorrectly identifying something as a true signature. Examples include: misidentifying a cat as a dog; claiming a file is infected, though it is clean; or categorizing someone as having cancer, when they do not.
- False Negative – Missing a true signature. Examples include: not identifying a picture of a dog as a dog; or failing to identify beer, and misidentifying it as olive oil.

Contrast these with the ideal case of a properly working ML identifier:

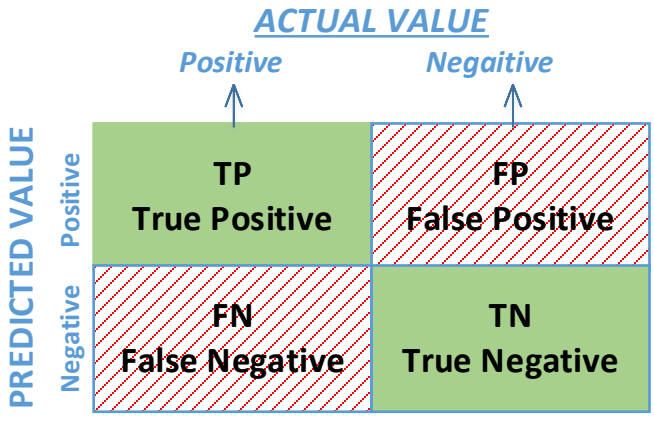
- True Positive – Correct identification of an actual signature/object
- True Negative – Correct identification of a false signature/object

In statistical data analysis, False Positives (FP) and False Negatives (FN) are also known as Type-I and Type-II errors. Note the interplay between the error types. Each case is unique, and the severity could depend on business objectives. A compromise might need to be reached between FP vs. FN detection. The ML engine can be trained accordingly.

Confusion Matrix

Confusion or error matrix compares the correct and incorrect predictions made by the machine learning classifier. The entries in the matrix show at a glance how the engine performs.

Table 2 – Confusion Matrix



With the above definitions, Accuracy becomes

Accuracy = true/ total = (TP + TN) /total

If the training sample batch is not evenly distributed, then the above formula will give skewed results.

Therefore other measures need to be considered.

Recall

Recall measures the prediction rate of true positives, out of all the correct predictions.

Recall = TP / (TP+FN)

Recall is a measure of the sensitivity of the ML classifier.

Precision

Precision is a measure of confidence of the positive predictions of the tool.

Precision = TP / (TP+FP)

The above measures can be adjusted by tuning the parameters of the ML engine. However, increasing one measure will decrease the other.

11.1. False Positives Example

In this example, the tool misidentifies the bright light in the dark background as fireworks (with a high confidence level).

Figure 2 - False Positive – Fireworks

Table 3 – Machine Learning Detection and Error Mitigation

Detected Category	Initial Confidence Level	New Confidence Level
Fireworks have been detected from 00:00:02 to 00:00:03	90%	< 30%

In the appendix a methodology to mitigate this issue is discussed. The third column indicates the updated confidence level after the mitigation is applied.

11.2. False Negatives Example

In this example, the tool fails to identify the alcoholic beverages in the image analysis. However, the term “cocktails” is noted in the audio transcript as depicted in the JSON file (Figure 3).

In the appendix a methodology to mitigate this issue is discussed.

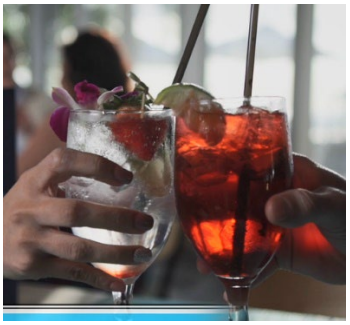


Figure 3 - False Negative - Alcoholic Beverage


```
{
        "id": 4,
        "text": "You can enjoy our hot tub cocktails and R Florida.",
        "confidence": 0.9069,
        "language": "en-US",
        "instances": [
         {
          "Start": "0:00:16.74",
          "End": "0:00:19.82",
         }
        ]
}
```


Figure 4 - JSON file of audio script of the parsed ad

The JSON file in Figure 4 indicates the word “cocktails” as parsed from the audio transcript. This data is available even though the image analysis failed to recognize that alcoholic beverages were in the video.

False Positives vs. False Negatives – Which is Worse?

The answer actually depends on the specific use case. The ML team needs to assess which is more critical -- missing a signature or tainted results?

The approach we recommend is to treat the former as critical. An example would be accidentally playing an alcohol or casino ad during a children’s program. But at the same time we recognize that a deluge of false positives would make the ML tool unpopular and the users would lose trust.

Another practical consideration is the nature of the output from the ML engine. Usually, it is a lengthy JSON formatted file containing a very large number of entries. Not all predicted values, though accurate, might be significant. In such a case, the calculated errors would be marginally small, leading to inflated accuracy claims. A clear methodology needs to be established before performing the error analysis. A rule-of-thumb would be to count only the events relevant to the task and above a pre-defined threshold value (e.g. 60%).

11.3. Limitations of Current Machine Learning Tools

To improve the detection accuracy, machine learning tools tend to use increasingly sophisticated algorithms. However, the algorithmic approach alone did not seem to produce expected results. Obtaining optimal results within a reasonable time is a challenge. Searching each video frame for a multitude of categories (alcohol, gambling, drugs, violence, trademarks, copyrighted content, explicit content, political content etc.) is time consuming. It could also be irrelevant (i.e. searching for all manners of firearms or medications within a beer ad would be wasteful).

The appendix section details a method to add a software engine to the workflow to perform additional analyses to enhance the results.

Appendix A Machine Learning Tool Requirements for Advertising Use Cases

A.1 Overview

Below is a collection of optional requirements specific to MVPD advertising applications. Network operators and service providers *may* use these for solution development and RFC preparation. The vendor partners might also find these helpful in developing product specs for carrier-class machine learning solutions.

For the initial phase the following limited work-scope could be considered.

- **Ad Ingest QC** – Identify restricted ad content per defined criteria. Develop interfaces to insert/integrate ML engine to the current workflow. The annotated results in the JSON file can be used to generate human readable reports.

- **Ad Classification** – Scan ads in the ad repository and generate metadata for ad cataloging and search. The search criteria and results display mechanism are discussed below.

See Figure 5 and Figure 6 for workflows.

A.2 General Requirements

a) Auto-detect and tag the media content by analyzing Video/Audio/Text components.
b) Identify following signatures and supply content descriptors with timestamps as applicable:
   b.1. People, animals and other general objects appearing in the media
   b.2. Provision to add user-defined objects/logos/emblems for the search criteria
   b.3. Sentiment and emotions, underlying topics as well as any anomalies detected
   b.4. Activities happening in the media
   b.5. Copyrighted content and popular trademarks
   b.6. Word phrases, sounds and other audio content that describe the media
   b.7. Length of the video content (e.g. 30 sec ad)

b.8. Language(s) spoken/displayed

c) Create a searchable index of the metadata derived by Machine Learning (ML) tool.
d) Provide a dashboard with user friendly UI for the Ops staff to search content metadata.
e) Create a workflow to seamlessly integrate machine learning results and data analytics.
f) Support multiple content types and formats. Note that the content can be file-based, stored in an archive, or streaming media.

A.3 Ad Ingest QC – Identifying Non-Compliant Content

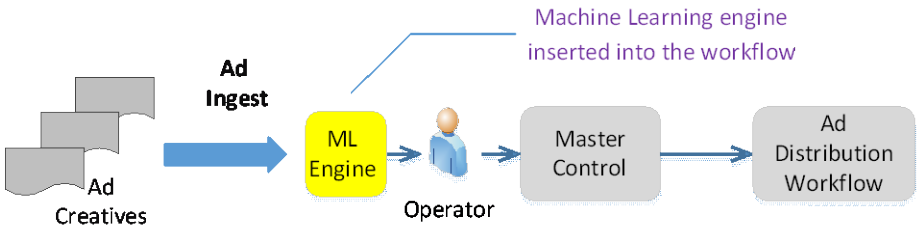


Figure 5 - ML tool usage in ad ingest QC and classification

a) Scan ads at ingest and generate descriptive metadata.
b) Create a summary of each ad content.
c) Tag the identified ads per IAB Content Taxonomy Mapping system: www.iab.com/guidelines/taxonomy . Use the hierarchical JSON format.

```
"26.3.3.7":{
    "Technology & Computing":{
      "Computing":{
          "Computer Software and Applications": "Digital Audio"          }
   }
}
```

d) Screen ads for restricted content. Flag the non-compliant ads for further processing: The table below shows restricted content in a typical (hypothetical) case.

Table 4 – Non-Compliant/ Restricted Content

Restricted Category	Requirement Importance	Comments/Examples
Alcohol	High	Beer, wine, hard liquor or variants, including those with non-descriptive names (“***Bud Light Lemon Tea***”).
Tobacco	High	Cigarettes, E-cigarettes, cigars, vaping
Drugs	High	Includes drug paraphernalia
Gambling, Casino	High	
Copyrighted content	High	Visual and/or audio track (songs, movie soundtracks)
Trademarked content	High	College brand T-shirts, logos and emblems

Restricted Category	Requirement Importance	Comments/Examples
Explicit Content	High	
Curse/Swear words	High	Profanity in ads is not allowed
Sexual products	High	
Violence	High	Guns, explosives, physical violence
Competitor content	Medium	
Political Content	Medium	(Hard to detect, but see below for plausible signatures*)

Note: For political content the following FEC requirement might apply: "...a "clearly readable" written statement that appears at the end of the communication, for a period of at least four seconds".* https://www.fec.gov/help-candidates-and-committees/making-disbursements/advertising/

e) Screen ads for quality issues (video jitter, macro blocking, audio clipping etc.)
f) Provide audio transcript of the media content emphasizing key words/phrases.

A.4 Ad Classification – Cataloging Ads in a Repository

a) Scan ads in the repository and generate descriptive metadata to classify the ads.
b) The metadata might be in the form of human readable "labels" in the UI/Dashboard.
c) The metadata might be in the form of JSON/text output files for programmatic analysis.
d) The metadata might summarize the ad content to enable cataloging.
e) The labels might be prioritized to better describe the ad content, as appropriate.
f) Use a search engine capability to parse ad metadata.
g) Given specified criteria, locate and retrieve matching ads from the ad repository.

A.5 Video Content Analysis Pertaining to Ads

a) Descriptive Metadata – Scan videos and generate descriptive metadata. Identify sentiments, underlying topics as well as any anomalies in the media content.
b) Searchable Catalog – Create a searchable catalog of videos based on tagged content.
c) Ad Recommender – Analyze video content and recommend ad opportunities for Contextual Advertising, including sentiment analysis.
d) Segmentation – Generate logical video segmentation boundaries and identify dominant activity (e.g. fight scenes, car chases, songs). Semantic and panoptic segmentation are advanced techniques that require additional processing.
e) Thematic advertising – Given an ad-campaign theme (e.g. eco-tourism), find matching videos from the collection. Find effectiveness of ads by different demographics/audiences.
f) Video Tagging – Screen video content for quality issues and tag accordingly.
g) Celebrities - Find videos of a given actor, including duration/time stamps, from a collection.
h) Closed captions – Translate speech to text for assets that currently do not have captions.

A.6 Reporting Requirements

a) Generate a pdf file containing the detected instances marked with screenshots, bounding boxes and time stamps.

b) The report might provide high-level and detailed-level data.

High Level – Aggregate and summarize the detections. Provide video, shot and frame level annotations. Provide audio transcript and summary of visually identified text (optical character recognition).

Detailed Level – Provide detections with counts (number of occurrences), durations or presence percentages (e.g. the occurrences per GoP (group of pictures) expressed as a percentage). Provide a detailed report of each identified signature (e.g. alcohol) along with thumbnail photos, start/end times, durations and confidence levels. The screen shots of interest might contain annotated bounding boxes to indicate the detections.

c) In addition to the above, a low-resolution version of the annotated video, with overlay bounding boxes is recommended (useful when a quick view of the detections is needed).

d) In addition to static reports, an active dashboard GUI with clickable labels (to indicate each occurrence on a time line).

A.7 Performance Requirements

a) Ability to process video content in near real-time (or within a specified delay).
b) Ability to process multiple files simultaneously.
c) Ability to prioritize ML job processing (beyond best effort/round robin modes).
d) Ability to meet specified latency requirements.
c) Ability to meet audio/video quality requirements, such as MOS, PSNR, SSIM as well as perceptual video quality.

A.8 Product Integration and Workflow Requirements

The machine learning classifier engines in the MVPD space are usually vendor developed and then customized for clients. One pitfall is that usually more emphasis is given to fine tuning the ML engine and less consideration is given to product integration. Both aspects are important.

A detailed assessment of ML module integration into the ad processing workflow is necessary prior to the production testing stage. Generally, this would cover interfaces, reliability/failover scenarios, data processing, performance and CDN integration. Figure 6 depicts this at high-level.

The output of the ML engine (audio, video and textual analyses results) is usually in JSON format. This raw data need to be converted to a more presentable format for human consumption. APIs also need to be configured for M2M communication with analytics engines, dashboards, etc.

a) **Report generation** – A pdf file containing the detected instances marked with screenshots and time stamps is recommended. A video overlay with bounding boxes is also desirable. An active dashboard GUI with clickable (hyper-linked) labels (depict each occurrence with time stamps) is a nice-to-have feature.

b) **Workflow integration** – Interfaces *should* be developed to insert/integrate ML engine to the existing workflows. A parallel process (vs. in-line), is recommended since the ML analysis speed is not close to real-time. The intent is to prevent any impact to the normal (non-ML) workflow functioning. While the workflows are generally in a single provider cloud, a hybrid-cloud scenario could also be envisioned (Figure 7).

c) **Interface development** – Interfaces *should* be developed to transfer JSON metadata to another module such as ad campaign manager or analytics engine.

API/Interfaces *should* be created as necessary for ad ingest and ad database classification scenarios:

- Auto-ingest feed of video clips to ML classifier engine
- Metadata output from ML engine to an ad campaign manager
- Ad Ingest – Metadata output from ML engine to GUI dashboard and other analytics systems
- Ad Classification Use Case – ML output metadata to be linked to an analytics/search system

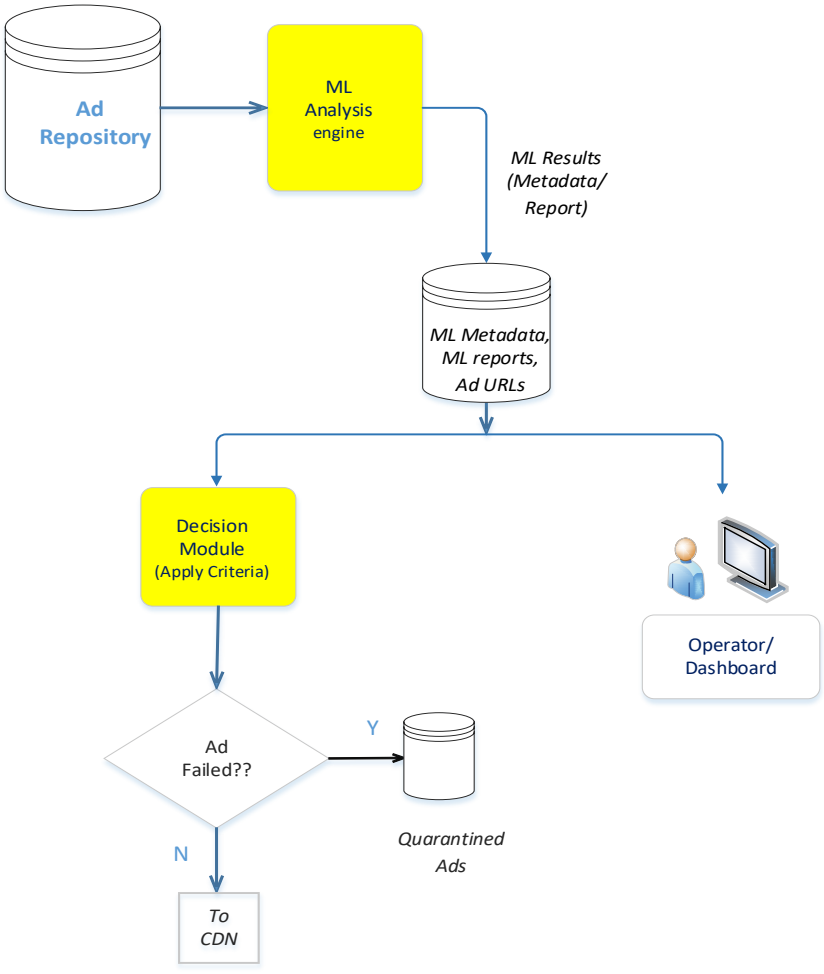


Figure 6 - Ad Classification Workflow

A.9 Cloud and API Requirements

a) Automated ML engine workflow integration *should* be configured for the following cloud scenarios:
   - Single cloud – Data resides on the cloud (or copied to). ML engine is located on the same provider cloud.
   - Hybrid (multi-cloud) – Data resides on the provider-1 cloud and copied to a different provider-2 cloud where the ML analysis is performed.

In each case the results are sent back to specified local servers.

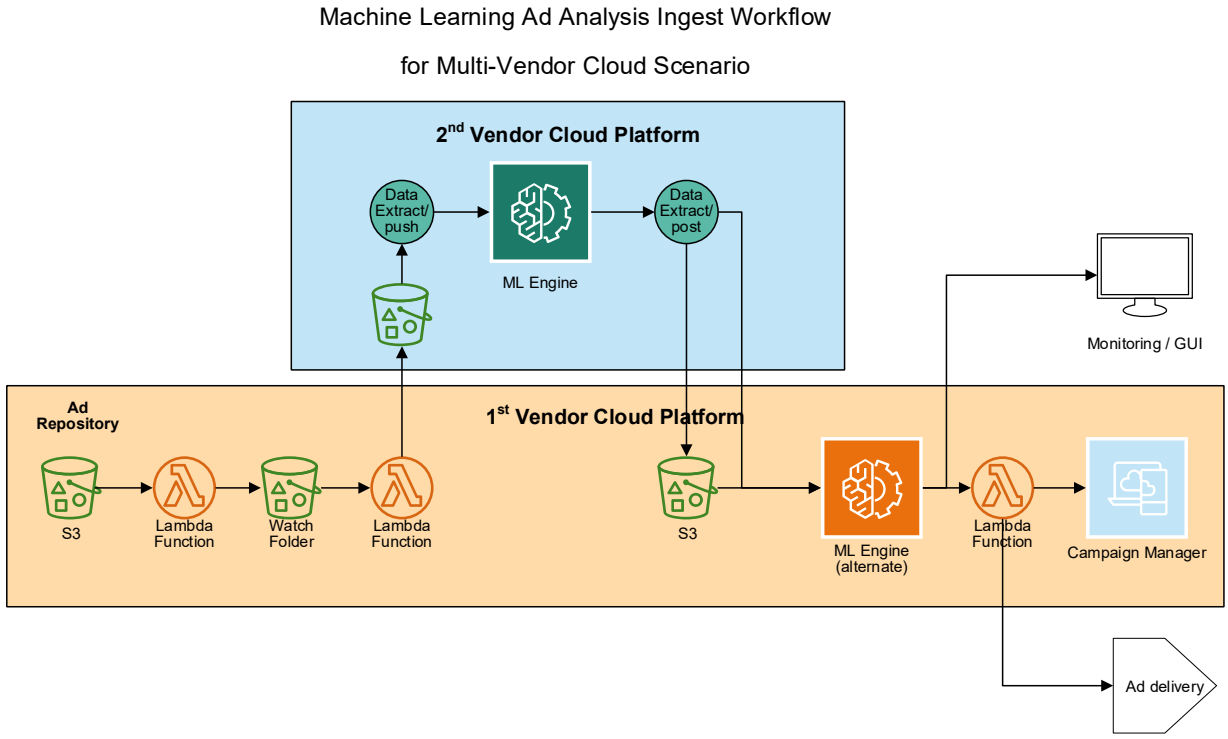


Figure 7- ML Ingest Workflow for Multi-Vendor Cloud analysis

(Note – The diagram shows AWS components for illustrative purposes, however it is applicable to any other cloud vendor product)

A.10 User Customization

a) It is important that the user is able to add custom objects to the ML analysis criteria. Examples would be product logos, new type of beer in the market, drug paraphernalia for detection.
b) If the tool has no provision for user customization, then the option for the modification to be effected via the vendor ML team.