A Project Report on

Content Sanitization Using CV and NLP

Submitted in partial fulfillment of the requirements for the award of the degree of

Bachelor of Engineering

in

Information Technology

by

Pranav Mayekar(19104040)

Under the Guidance of

Ms. Rujata Chaudhari
Ms. Shital Agrawal

Department of Information Technology
NBA Accredited
A.P. Shah Institute of Technology
G.B.Road,Kasarvadavli, Thane(W), Mumbai-400615
UNIVERSITY OF MUMBAI

Academic Year 2022-2023

Approval Sheet

This Project Report entitled ***"Content Sanitization Using CV and NLP"*** Submitted by ***"Pranav Mayekar"(19104040)***is approved for the partial fulfillment of the requirenment for the award of the degree of ***Bachelor of Engineering*** in ***Information Technology*** from ***University of Mumbai***.

(Ms. Rujata Chaudhari)
Guide

(Mrs. Shital Agrawal)
Co-Guide

Dr. Kiran Deshpande
Head Department of Information Technology

Place:A.P.Shah Institute of Technology, Thane
Date:

CERTIFICATE

This is to certify that the project Synopsis entitled ***"Content Sanitization using CV & NLP"*** Submitted by ***Pranav Mayekar (19104040)"*** for the partial fulfillment of the requirement for award of a degree ***Bachelor of Engineering*** in ***Information Technology.***to the University of Mumbai,is a bonafide work carried out during academic year 2022-2023

(Ms. Rujata Chaudhari)
Guide

(Mrs. Shital Agrawal)
Co-Guide

Dr. Kiran Deshpande
Head Department of Information Technology

Dr. Uttam D.Kolekar
Principal

External Examiner(s)

1.

2.

Place: A.P.Shah Institute of Technology, Thane
Date:

Acknowledgement

We have great pleasure in presenting the report on **Content Sanitization using CV and NLP** We take this opportunity to express our sincere thanks towards our guide **Ms. Rujata Chaudhari** & Co-Guide **Ms. Shital Agrawal** Department of Information Technology, APSIT thane for providing the technical guidelines and suggestions regarding line of work. We would like to express our gratitude towards her constant encouragement, support and guidance through the development of project.

We thank **Dr. Kiran B. Deshpande** Head of Department,Information Technology, APSIT for his encouragement during progress meeting and providing guidelines to write this report.

We thank **Prof. Sonal Jain** BE project co-ordinator, Department of Information Technology, APSIT for being encouraging throughout the course and for guidance.

We also thank the entire staff of APSIT for their invaluable help rendered during the course of this work. We wish to express our deep gratitude towards all our colleagues of APSIT for their encouragement.

Pranav Mayekar (19104040)

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, We have adequately cited and referenced the original sources. We also declare that We have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

_____

Pranav Mayekar (19104040)

Date:

Abstract

We have created a platform named coinplanet where new projects and new businesses can collect information to give a headstart to their projects with the help of surveys and detailed analysis of responses with support for multimedia data like image, video and audio. As the surveys sometimes tend to be anonymous they are subject to online menance where the survey responder might put any inappropriate text or multimedia data. However with increasing number of people using the internet they might find image, videos that they might not want to watch. Most of the people do not want to witness violent acts, such as one man slashing another man with a knife or nude scenes. However, a person can wind up seeing it on the platform because these kinds of sequences are becoming a typical occurrence nowadays. By using artificial intelligence and machine learning, the platform will be able to regulate the videos and other multimedia data that the users consume. In addition, due to individual preferences, some people might not want to witness a nude scene, an adult scene, or a murder scene. Consequently, it is useful to them as well. Also the growth in technology has given rise to many businesses of various sizes. With the increase in number of businesses there is more demand for online platforms to serve their different needs and to outsource their tasks. As the usage of online services is increasing it has also given rise to wide variety of online menace like hate speech, toxicity, insult, obscenity etc. To deal with this we need sophisticated mechanisms to handle this kind of data and to prevent them from spreading further to keep the online space safe. This can be done with the latest technologies like Artificial Intelligence and Machine Learning. There is also a need to handle not only text data but also multimedia data like image, video, audio. The users on the platform are given the liberty to either safely view the multimedia data or block the toxic data entirely. It also enriches user engagement towards our website.

Contents

List of Figures

List of Abbreviations

AI:	Artificial Intelligence
ML:	Machine Learning
CV:	Computer Vision
NLP:	Natural Language Processing
CNN:	Convolutional Neural Network
ConvNet:	Convolutional Network
IT Act 2000:	Information Technology Act 2000
SVM:	Support Vector Machine
VGG:	Visual Geometry Group
UI:	User Interface
RabbitMQ:	Rabbit Message Queue
NLU:	Natural Langugage Understanding

Chapter 1

Introduction

In Today's developing world every individual is trying to set up their own businesses or Start a new project for which they will be recognised for their work. So the entrepreneurs or the new establishments who are trying to make their own place in the market face a lot of difficulties to stand out as a unique business. As new establishments and businesses need a good amount of data about their target audience, the market segments, information about their competitors etc, to lay their ground work, and to gather such huge information the businesses spend a lot of capital to test and analyse their products.

For newly established businesses and new projects we have created coinplanet where businesses and individuals can gather unbiased opinions and information through surveys to collect and analyze information. Businesses collect feedback from their clients who are sharing their views about a particular product, however some users might use inappropriate text or data while giving feedback. To prevent inappropriate content from spreading on the platform content sanitization comes in to frame by detecting any inappropriate text/data.

As coinplanet also supports the framework where a user can also create a survey by adding various multimedia like images, videos and audios. So coinplanet will sanitise that data as well, before the survey is published. It will sanitise the data to ensure that the survey does not content any inappropriate content and detect to detect any usage of foul language as well so that if a user tries to enter such data in the survey the system will automatically detect and discard those responses to help the owner of the surveys to get good and quality data.

To carry out the implementation of such techniques we are using Machine Learning Technologies. We will be classifying the multimedia in different categories to calculate their toxicity and based on that the system will analyse and detect the inappropriate content. For Image we will be analysing factors such as texture, contrast and colour based factors and for video we will carry out the same procedure by analysing the images frame by frame as for audio we will be using NLP.

The nudity detection part is implemented with the combined use of two libraries of the CNN algorithm. The first library i.e. Open CV, mainly detects human figures from the images and the fragmented frames for video and crop them out in a separate image file. Then CNN is used on the new image files to determine if the images are nude or not. The threshold that has been currently set in the prototype is 60 percentage. Any frame that has a nudity percentage of over 60 percentage is considered as an inappropriate scene and is therefore eligible for blurring. We aim to train the model over inappropriate scenes from the internet to make the process as accurate as possible.

An artificial neural network called a convolutional neural network, often known as a CNN or ConvNet, has so far been most frequently employed to analyse images. CNN is quite helpful for image analysis because it has some kind of specialised mechanism for being able to pick out or recognise patterns. It is reported to operate on a topology resembling a grid. In addition to providing the vision for robots and self-driving cars, CNN has proven successful in recognising faces, objects, and traffic signs. CNN has a good success rate for identifying several things.

1.1 Purpose

The purpose of Content Sanitization is to ensure the safety and privacy of users, as well as to protect computer systems and networks from potential security threats. In addition to security concerns, text sanitization can also be used to ensure that text data is clean and consistent, making it easier to process and analyze. This can be particularly important in natural language processing (NLP) applications, where text data is often used to train machine learning models. Content Sanitization will help filter out or block content that is deemed obscene, offensive, or inappropriate. The purpose of this platform is to protect users, particularly children and other vulnerable populations, from exposure to harmful or offensive content. By filtering out inappropriate content, it will help maintain a safe and appropriate online environment. In some cases, it may be used to protect national security or public safety, such as by suppressing information that could compromise military or intelligence operations, or by blocking access to content that incites violence or hate speech.

1.2 Problem Definition

1.2.1 Problem Identified:

Our website coin planet helps companies and individuals to create survey forms to collect feedback from their clients who are sharing their views about a particular product, however some users might use inappropriate text or data while giving feedback In our platform a user can create these survey forms and add contents like text, image, video and audio in those survey forms. The only issue is that if a user adds some inappropriate contents like obscene images and toxic language in the survey forms and then publishes the survey. Users of varying age and from different backgrounds visit the platform everyday, it is our duty to make sure that the content on our platform is safe for everyone who visits our platform. Different people from all over the world have their own belief systems. Also if the toxic content is not managed then the government has laws mentioned under IT Act, 2000, Section 67 under which severe punishment is given to individuals or organizations not following the guidelines.

1.2.2 Solution Proposed:

To prevent inappropriate content from spreading on the platform content sanitization comes in to frame by detecting any inappropriate text/data. It will sanitise the data to ensure that the survey does not content any inappropriate content and detect to detect any usage of foul language as well so that if a user tries to enter such data in the survey the system will automatically detect and discard those responses to help the owner of the surveys to get good and quality data

1.3 Objectives

We are implementing this project to meet the following objectives.

- To manage toxic text by categorizing it into 6 categories that are Toxic, Obscene, Threat, Insult, Severe Toxic, Identity-hate using Logistic Regression Model for each of the category.
- To detect and manage inappropriate images using Computer Vision by making use of Convolutional Neural Network to detect features necessary for inappropriate image detection.
- To perform feature extraction and regulate inappropriate audio by converting language to text using Natural Language Processing and to beep inappropriate words in the audio.
- To detect and manage inappropriate scenes in a video by using Computer Vision and Convolution Neural Network to automatically blur the frames of the video.

1.4 Scope

- It can be used for censorship of films and videos, like it will automatically blur the images which contains obscene scenes and nudity.
- Audio censorship it will beep the words where abusive language or swear words are used and will provide you a censored audio.
- It can be used by government organizations to reduce toxicity while conducting surveys or feedback.
- It can be used by companies to automatically sanitize the data by using our obscenity blocker as a solution.
- Prevent havoc on any online platform by controlling the toxicity.
- Handle any amount of traffic with this scalable obscenity blocker solution.

Chapter 2

Literature Review

- According to Clayton Santos, Eulanda M and Eduardo Souto [1], image zoning works best for identifying the inappropriate features in the image. They have observed that instead of performing feature extraction directly on the entire image, they divide the images into zones which are then used for local feature extraction. So they presented a strategy that uses image zoning along with the analysis of six features, two color –based and four texture-based features, applied to nudity detection in images.In order to illustrate the whole nudity detection process, they have adopted the architecture most often applied in works that perform skin filtering. First, input images are normalized and then segmented through a filtering mechanism. Then feature extraction is performed by dividing images in zones. Thus, features obtained from each zone are combined into a feature vector and submitted to SVM for classification on nude or neutral (non-nude) classes.
  Their study also confirms the hypothesis s that nude images have most of its skin pixels concentrated on the central region of the image. Experiments conducted using skin filtering algorithm and zoning strategy demonstrated that features extracted from the central zone (first zone) provide the most relevant information for SVM classifier. Besides, the zoning strategy increased recognition rates effectively.

- According to Dipta Roy, Raisa Siddiqui, Rahat shahriar Islam [2], openCV is the best library to process videos so they have proposed a system where the first step they have implemented is to convert the video file into its individual frames. In order to do that, they used Open CV(Open Source Computer Vision Library) to convert the video into frames. Open CV is an open source computer vision and machine learning software library. Open CV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products. The number of frames will be fetched from the directory where the frames are stored and then they have applied a logic where if n is less than the number of frames then it will read the frames of the video and save it to directory with frame number. Unless the condition becomes false the procedure will keep on running. we have also used the above mentioned procedure as well to convert video into frames by using Open CV.

- According to the study made by Shoji Kido, Yasusi Hirano, Noriaki Hashimoto [3], alexnet was the best CNN implemented by that time so they have used alexnet to train over million images. They used a CNN as a non-feature extraction based approach. In this study, they used relatively small number of image cases. Therefore, they used the pre-trained CNN model "AlexNet" which has been trained on the ImageNet dataset, which has 1,000 object categories and 1.2 million training images [3, 4]. They also used data augmentation for supplementing a small number of image data. In the first step, image data were augmented by 8 times with rotation and reflection.For lung nodule cases, they have used 163 nodules, and we obtained 1,304 nodules with data augmentation.For diffuse lung disease cases, they have used 9,635 patches (715 CON, 1,051 HON, 1,886 GGO, 3,474 EMP, and 2,509 NOR) from 372 patients, and obtained 77,080 patches with data augmentation.

  In all cases, same number of input images for benign or malignant nodules or all patterns of diffuse lung diseases were randomly selected. The images were divided into training and validation sets. 30% of images from each set were selected for the training data and the remainder, 70%, for the validation data. Randomization was performed to avoid biasing the results. The CNN model processed the training and test sets. In the first, features of training images were extracted using CNN. Next, these features were used to train a multiclass support vector machine (SVM). Finally, the multiclass SVM classifier evaluated validation set.

- According to the study made by Bulbul Bamne, Neha Shrivastava, Lokesh Parashar and Upendra Singh [4], there was a need to improve the pre-existing convolutional neural network. In this research paper they have improved pre-existing convolutional neural network using 2 techniques that are transfer-learning and majority voting ensemble method. The convolutional neural network comes with different layers that are convolution layer, pooling layer and fully connected layer. The very first layer consists of filters which deal with the width and height of the input, it also detects shapes, brightness, edges etc for the forward system. In this paper, the initialisation of the weights for filter is made randomly, and then the iteration is performed to update the weights during the training process. The discretisation of the samples is done in the pooling layer. This is done to reduce the dimensions of the input data. The fully connected layer performs the classification. The first method is transfer-learning where pre-trained convolutional neural network is retrained using some other dataset by using previous weights. Some popular pre-trained CNN models are VGG19, VGG16, AlexNet, GoogleNet, ResNet. The second method is majority voting where N Classifiers are trained on the same training dataset. The N classifiers are then used to make predictions using the new test dataset. Classifier 1 makes the prediction y head 1, classifier 2 makes the prediction y head 2 and so on. Then Majority Voting is done where N predictions are converted to 1 prediction. We have used the second procedure from the above comparison in this project.

- As per the study made by Midia Yousefi and Dimitra Emmanouilidou [5], toxicity seems to manifest not just locally, but throughout a phrase/sentence, therefore they created a mechanism which summarizes the frame-level feature map into an utterance-level feature vector. They have proposed a self-attentive CNN Architecture to detect toxic speech based on acoustical features. They have carried out these toxicity classification in two ways extracting features mostly representative of toxic samples classifying them into toxic or non-toxic content. In the first step they are using a CNN architecture that learns higher level information from the spectral features of speech. To tackle this toxicity situation they implemented two alternate e attention mechanisms called "Learnable Query Attention" and "Self-Attention". In learnable query attention the core idea of attention is to compress all the important information of a sequence into a fixed-length vector, so that computational resources can focus on a restricted set of important elements and in Self-Attention the idea is to tackle the challenge of learning a universally robust Query and to have abstractive summarization and image description generation as well.

Chapter 3

Proposed System

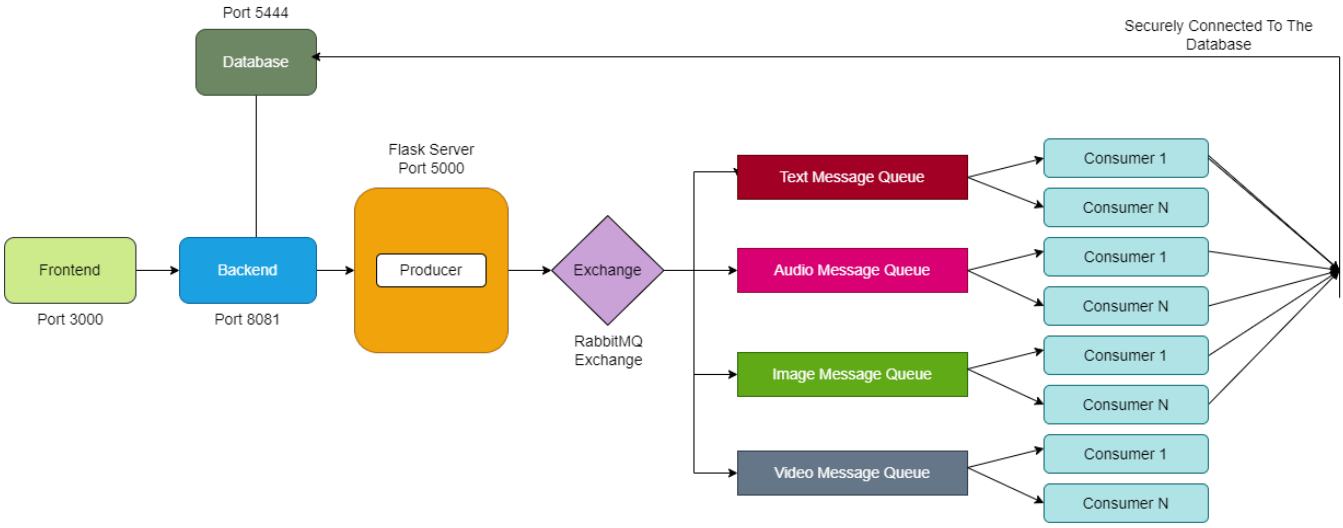

Figure 3.1: Proposed System Architecture

The flow starts from the Frontend like any normal request-response model. The Manage Survey Page has a section called Preview where one can see the state of the survey if is published or not. If it is to be published then the User needs to click on Publish Button on the UI. On clicking the Publish button the request is sent to the backend.

The backend processes the request to see if it is valid or not using the operational data from the database. If it not valid then it is rejected and the flow stops there itself.

However if it is valid then the request is sent to the Flask Server running on Port 5000 where the Producer code resides. On hitting that particular url with the surveyId as the body of the message the message is the forwarded by the Producer to the Exchange.

The Exchange is the main box where the routing of messages is present as defined by the user. It decides to which message queue the message should be sent. Here only 1 message is present so that message is sent to that queue.

Here 3 consumers are present which behave in a competing consumer pattern to consume the message from the message queue.

The consumers contain the actual code for running the Machine Learning Models and do the required prediction to see if the text data is toxic or not. If some of the text is toxic then the survey is marked as toxic and therefore the survey cannot be published.

The flow is similar for responses of a survey. When the request is sent to the Flask Server running on Port 5000. There is a specific url where the producer code resides for sending "surveyId" as well as "responseId" in one message as the body of the message to the Exchange. The exchange will route the message to the specific consumers where the code for handling toxic text, image, video, audio in the response will be present.

When the consumer code first starts executing it will connect with the database and obtain all the text, image, video, audio corresponding to that response of the survey according to the "responseId". The consumer code makes use of the models created for handling image, audio and video. The survey owner has the option to either discard the toxic images, videos, audios entirely or to safely and automatically handle them in the following way by blurring the image, beeping the audio when toxic audio is encountered, blurring the video frame when toxic frames are encountered.

We have used different machine learning algorithms for sanitizing different multimedia contents so that our system will carry out separate analysis of each multimedia content and give us a summarized result on whether the content is inappropriate or toxic. We have integrated different models by creating separate consumer packages using Rabbit MQ as the consumer is the one that contains the Machine Learning package and the Consumer package uses a function from that Machine Learning package to predict the results.

Text Sanitization:

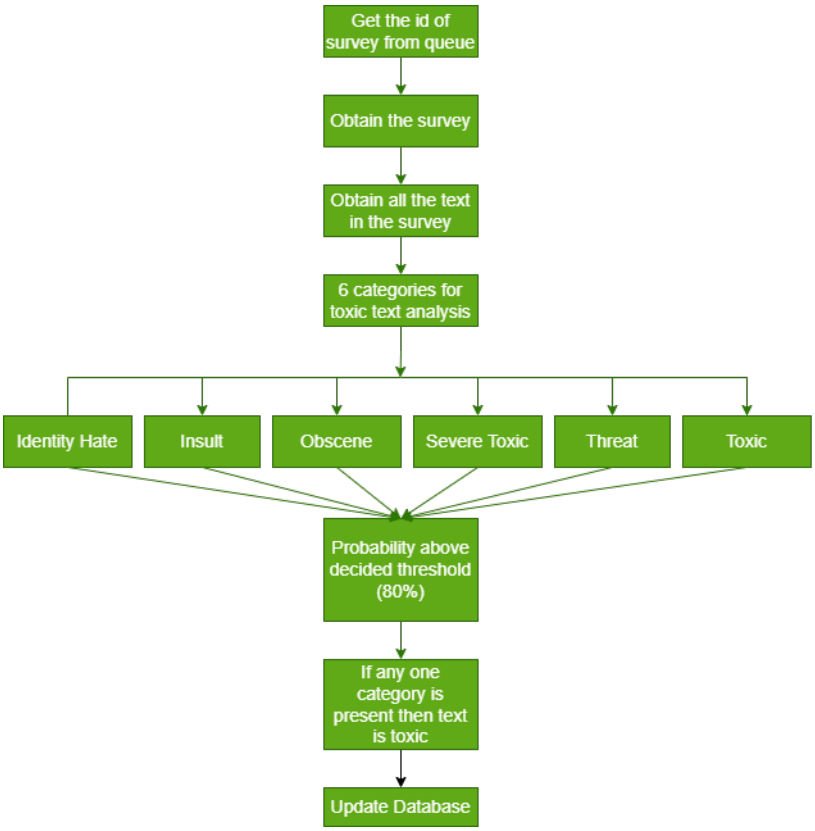

Figure 3.2: Working For Text Sanitization

We have used logistic regression as it is an useful analysis method for classification problems. This model will helps us predict the probability as to how much extent is the text toxic. we have created six different categories here namely Toxic, Obscene, Threat, Insult, Severe Toxic, Identity. We are using the Model for these 6 categories separately and if any of the category results in a probability higher than 0. or 80% then that survey will not be published unless and until the text is changed and the user publishes the survey again.

According to the need each category can be given different threshold. Like for example the category Obscene can be given a lower threshold like 60% so that as soon as little obscenity is detected in the text it will mark it as toxic. On the other hand category like Insult can be given a relatively higher threshold like 90% so that it will mark the text as insult only if the text is strongly an insult statement.

Image Sanitization:

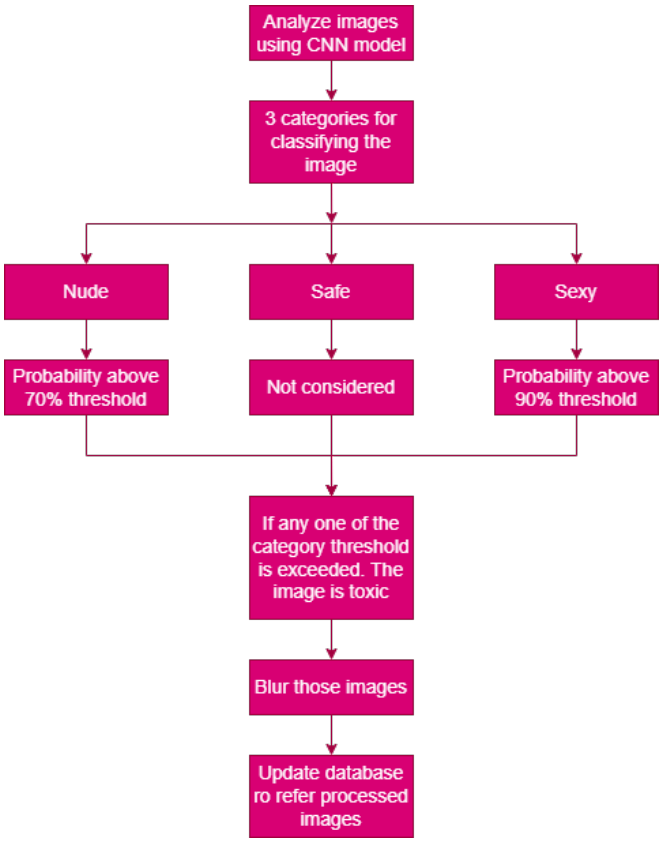

Figure 3.3: Working For Image Sanitization

For image sanitization our main focus is on nudity detection. we have used CNN for analysing, processing and classifying the images on six classification factors among which two are color based and four are texture based using the image zoning strategy. Firstly we are normalizing the input images and then the images are segmented through a filtering mechanism. After that the entire image is divided into different zones to extract the features locally of the image. Thus, features obtained from each zone are combined into a feature vector and by passing them through various convolution layers and pooling layers then we determine the classification on nude or neutral (non-nude) classes.

Working for Video Sanitization:

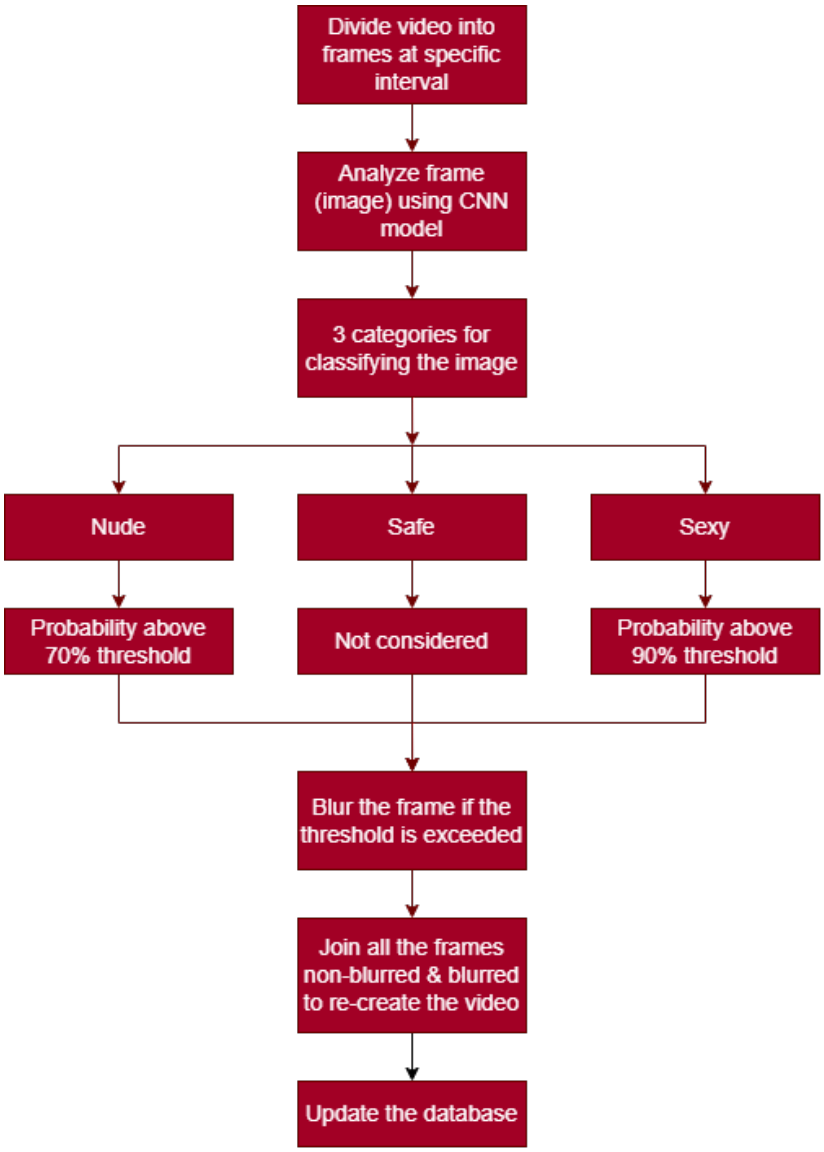

Figure 3.4: Working For Video Sanitization

The same concept of image sanitization is implemented over here, the only thing is the whole video is divided into multiple frames and each frame is analysed using image zoning , skin filtering and feature extraction strategies. To convert the video file into individual frames we have used Open CV. The number of frames will be fetched from the directory where the frames are stored and then we applied a logic where if n is less than the number of frames then it will read the frames of the video and save it to directory with frame number. Unless the condition becomes false the procedure will keep on running. so that we get to know that if the video is toxic or inappropriate.

Working for Audio Sanitization:

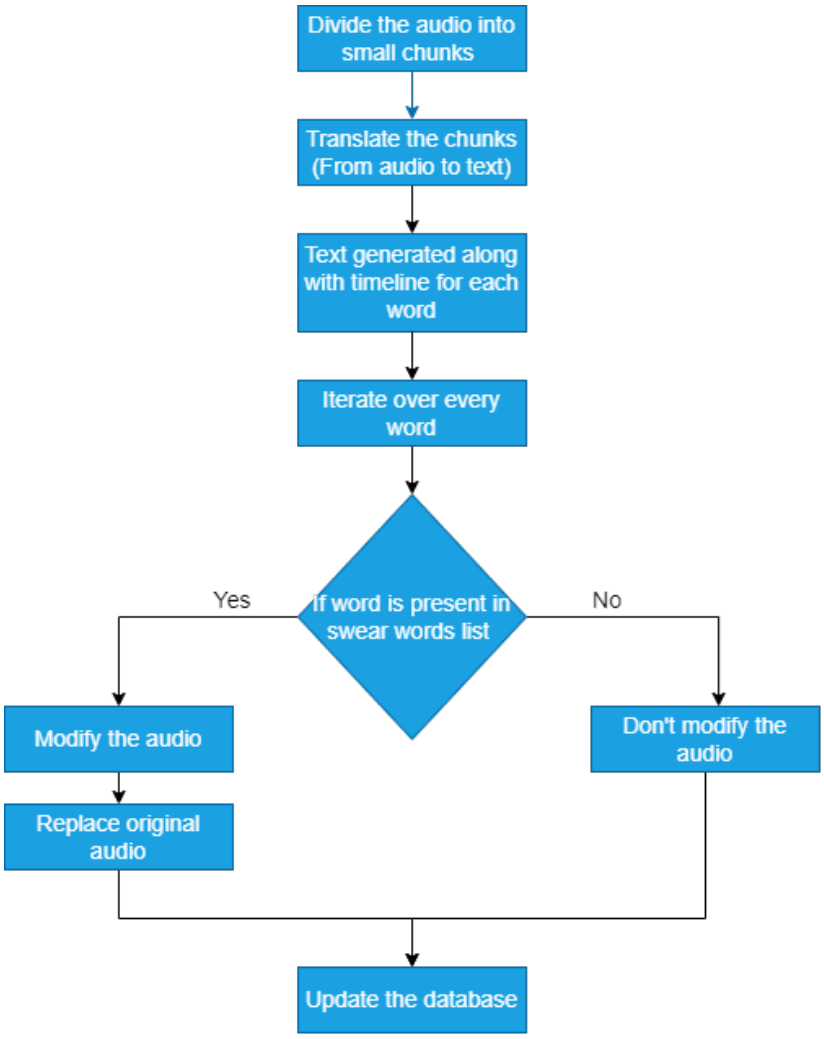

Figure 3.5: Working For Audio Sanitization

For Audio Sanitization we have used Natural Language Processing (NLP). Where the audio file is analysed and using 'NLU' we are converting the audio into text format and then we are implementing text sanitization where we have used logistic regression. The audio is divided into several chunks and by using NLP models we are looking for any inappropriate or swear words. After detecting the words we beep that part and the rejoin the chunks again. so that the audio is properly processed while submitting the form again.

Chapter 4

Project Design

Use diagrams describe the high-level functions and scope of a system. These diagrams also identify the interactions between the system and its actors. The use cases and actors in use-case diagrams describe what the system does and how the actors use it. It is best for describing our main functionality of survey creation and toxic content management.

Toxic content sanitization comes into play when the request reaches the backend and the backend then queues the survey for sanitization process. Once the message is popped out from the message queue it is consumed by the consumers. The consumers then hold the logic to perform analysis on the data according to the type of consumers. Following are the types of consumers responsible for sanitizing specific data.

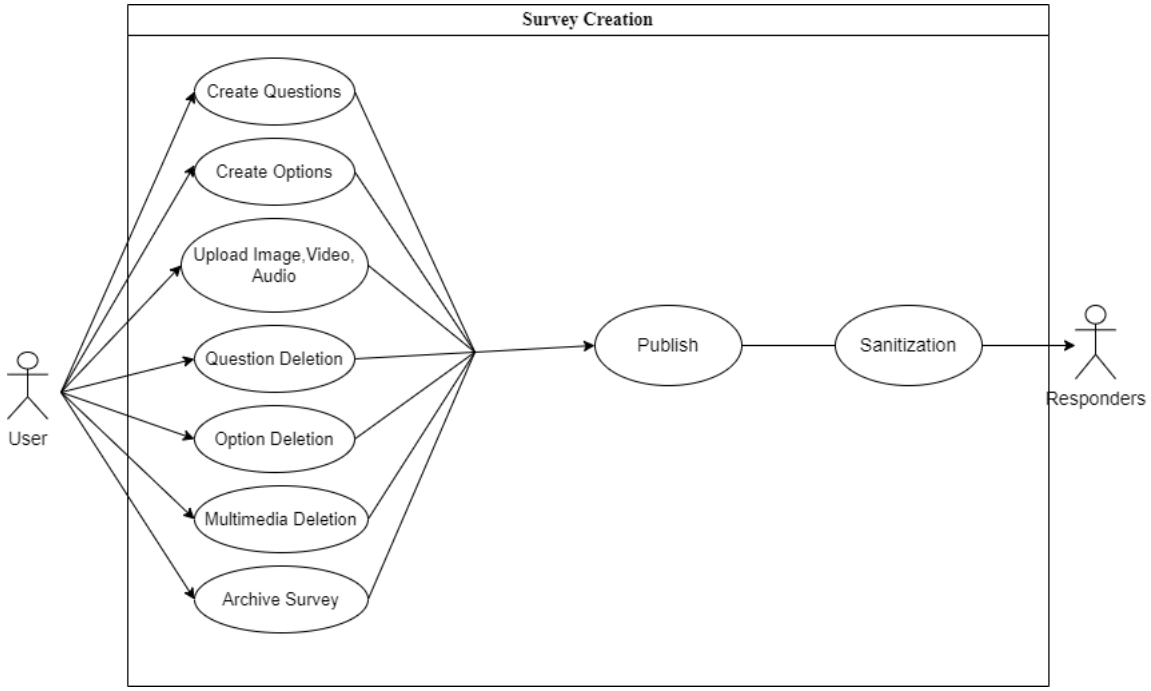

Figure 4.1: Use case Diagram

4.1 Text Sanitization Consumer

When the user filling out the form enters some text in form it is saved in the local state and on submitting it is ultimately sent to the consumer. The consumer analyzes the text sentence by sentence if it belongs to any of the 6 toxic categories it is marked as toxic and alert is shown to the user. And the form cannot be submitted until the user removes the toxic text.

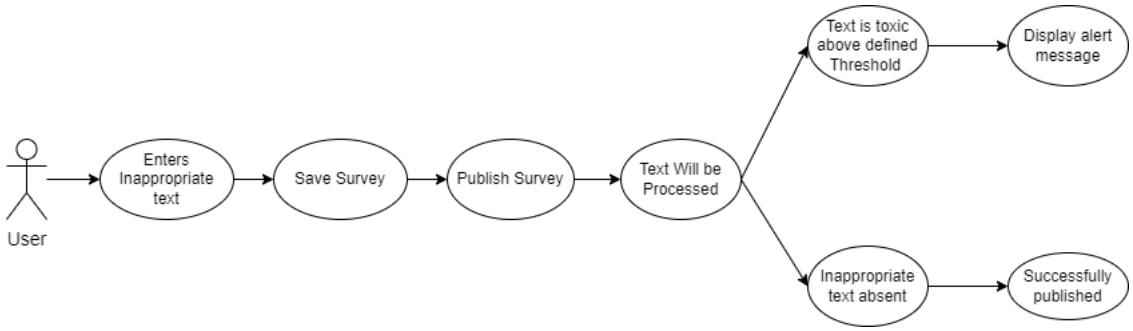

Figure 4.2: Use Case Diagram for Text

4.2 Image Sanitization Consumer

Image sanitization comes into play when the user uploads an image. The image is sent to the image processing consumer. The image is analyzed by using CNN model which classifies the image into 3 categories Nude, Safe, Sexy. Each category has a different threshold for classifying the image as toxic. If the image is toxic then it is blurred and the original image url is replaced with this new url pointing to the blurred image.

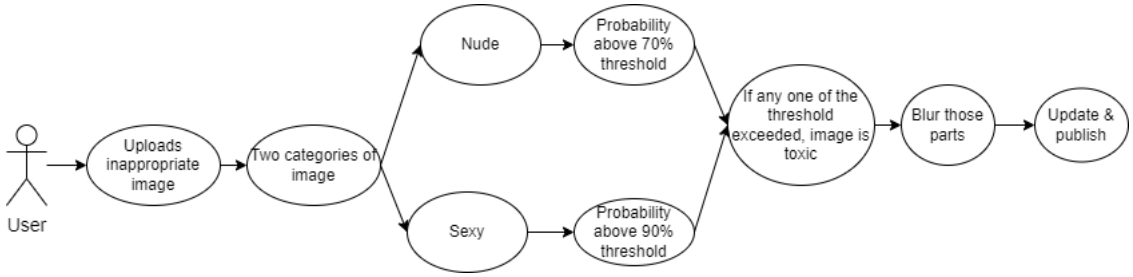

Figure 4.3: Use Case Diagram for Image

4.3 Video Sanitization Consumer

As this survey supports video files which can be viewed directly in the form. The form is susceptible to toxic videos where some inappropriate scene might be present. For handling the inappropriate video, the video is first divided into frames and the frame which is basically an image is analyzed using the same CNN model used for images. After blurring the images all the frames are patched together. However during this process the audio is removed automatically. The original audio is applied over the video and the original video url is replaced with the new video url pointing to the new video.

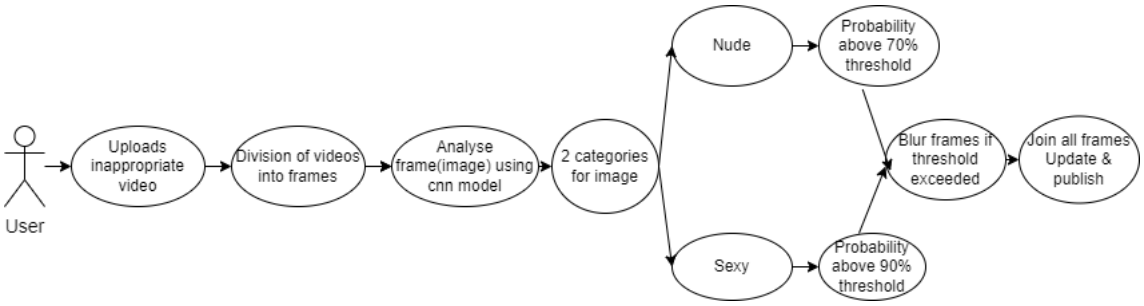

Figure 4.4: Use Case Diagram For Video

4.4 Audio Sanitization Consumer

Audio sanitization comes into picture when the audio has been uploaded from the device. Audio uploaded by using url from different domain is skipped for processing as the audio doesn't belong to this platform. The original audio is divided into chunks. Each chunk contains is then translated to text by using vosk model. This gives us the text with it's corresponding timestamp. If any swear word is present in the audio it is automatically beeped. Then these chunks are stitched together recreating the audio with beeped words. The original audio url is then replaced with the new audio url pointing to the new audio with specific beeped words.

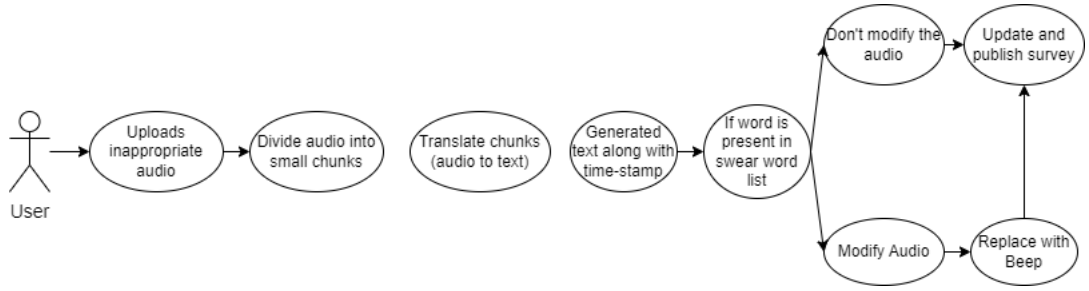

Figure 4.5: Use Case Diagram For Audio

Chapter 5

Project Implementation

The project implementation for handling toxic data like text, image, video audio is shown below. When the survey is first created the request for sanitization is first sent to the backend flask server which is seperate from the backend for handling business logic. This flask server consists of producer code from where the surveyId is encoded as a message and sent to the producer. Using pika library for interacting with the RabbitMQ.

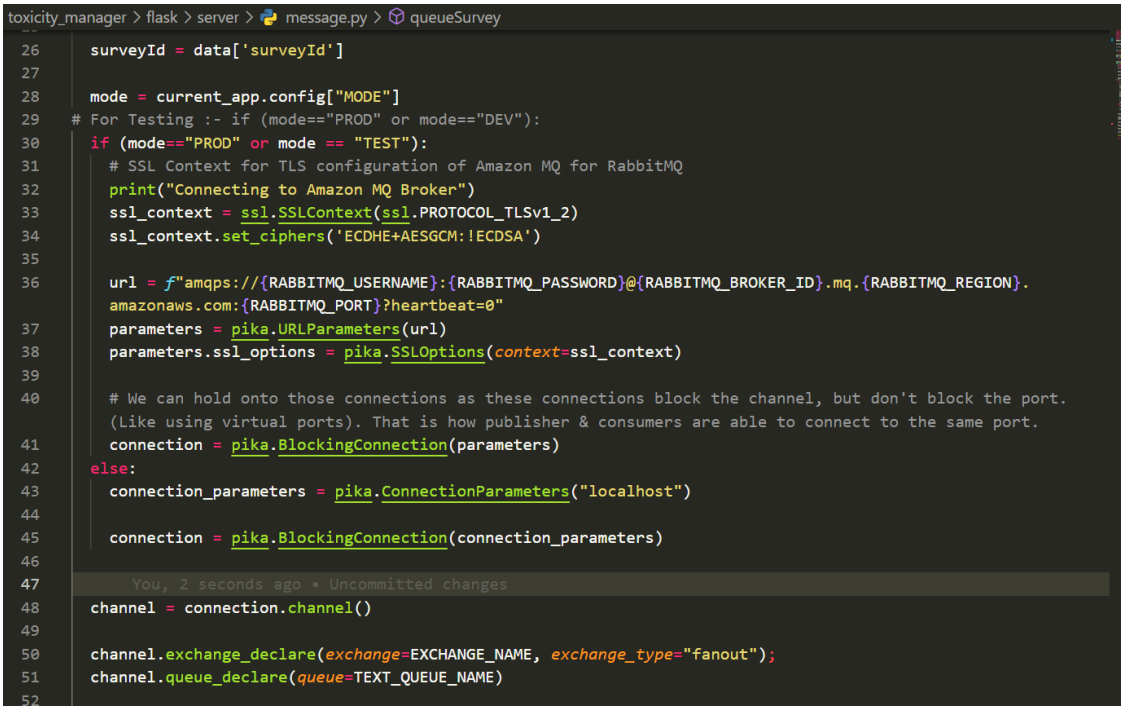

Figure 5.1: AI/ML Server Implementation

After the exchange receives the message that is surveyId it is forwared to all the connected message queues. Because the exchange is a fanout exchange it means all the connected queues to the exchange will receive the message sent to the exchange.

The connected queues to the exchange forwards the message received from the exchange to it's corresponding consumers. Consumers consume the message and process the survey text sentence by sentence. It is responsible for loading the six logistic regression models

which gives the probability if the sentence belongs to the following categories identity hate, insult, obscene, severe toxic, threat, toxic.

Figure 5.2: Text Consumer Implementation

The audio consumer is responsible for sanitizing the audio uploaded by the user from the device and not via any url. The audio consumer receives the message that is surveyId from the exchange and it firsts analyze the audio by breaking the audio into chunks. Afterwards the chunks are analyzed to convert audio to text. Afterwards the text is analyzed for swear words. After collecting the timestamps of such words the chunks are stitched together with the toxic words automatically beeped.

Figure 5.3: Audio Consumer Implementation 1

The below image shows how the chunk is divided into words along with timeline which is further used for beeping.

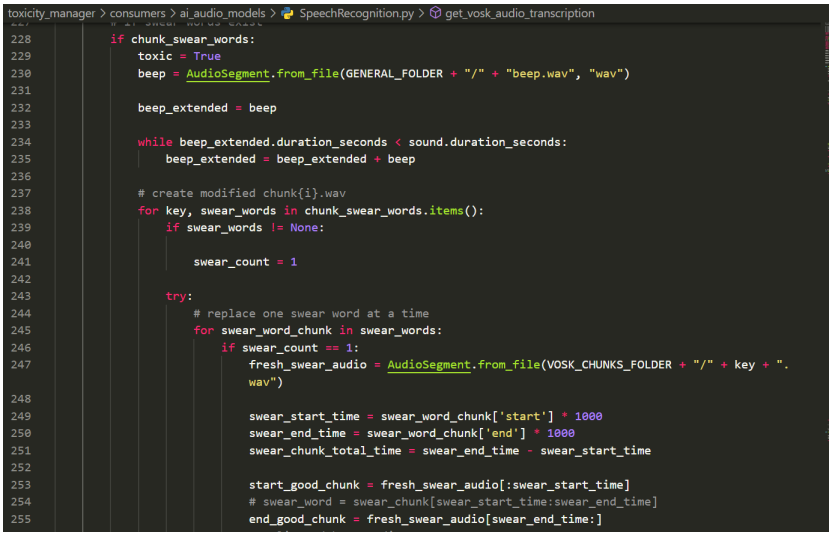

Figure 5.4: Audio Consumer Implementation 2

The image consumer is responsible for sanitizing the image uploaded by the user from the device and not via any url. The image consumer also receives surveyId from the fanout action performed by the exchange and from it's attached queue. The image model is a Convolutional Neural Network trained using Transfer Learning. The model gives the probability as to the image belongs to which category among nude, safe, sexy and to what extent.

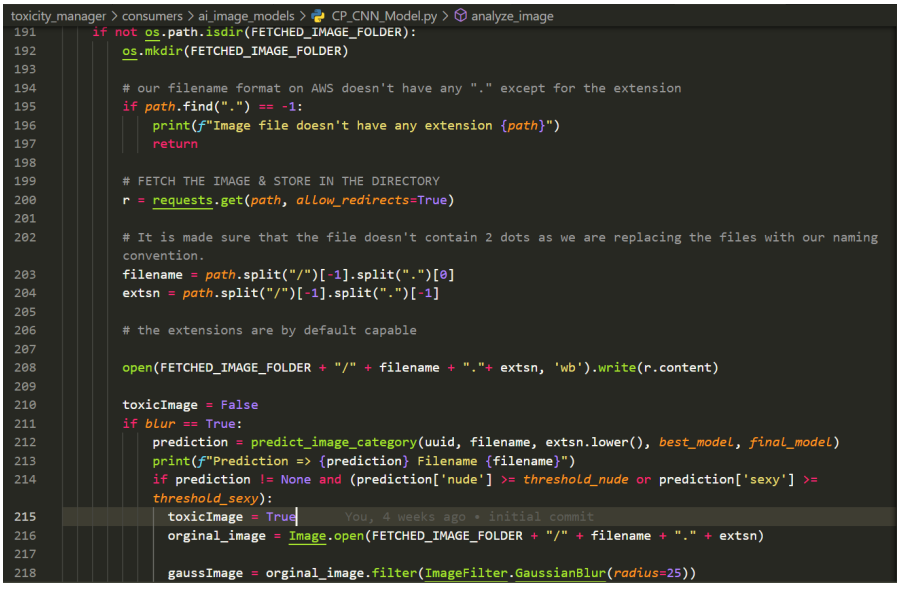

Figure 5.5: Image Consumer Implementation

The video consumers consumes the surveyId from the video message queue. After obtaining the survey it obtains the corresponding videos in the survey. The videos uploaded by user from the device are analyzed and the video processing happens in the following way. The video is fetched and video is converted to images by obtaining each frame. Each frame is analyzed and if it contains any toxic frame showing nudity, obscenity or any inappropriate data it will be blurred using gaussian blur.

Figure 5.6: Video Consumer Implementation 1

Figure 5.7: Video Consumer Implementation 2

Chapter 6

Testing

Software testing is the process of evaluating and verifying that a software product or application does what it is supposed to do. The benefits of testing include preventing bugs, reducing development costs and improving performance.A test management plan helps to prioritize which types of testing provide the most value – given available time and resources.

6.1 Functional Testing

Functional testing is an activity that extends from component-level testing all the way to regression testing of existing features. In our project we have performed various types of functional testing namely component testing, integration testing, end to end testing

6.1.1 Component Testing

In this method, tester selects a function and gives input value to examine its functionality, and checks whether the function is giving expected output or not. Component testing is similar to unit testing because they both isolate a single functionality and validate that individually. However, testing for components at this phase might call for stimulative interactions with sample test data, aka stub and driver.

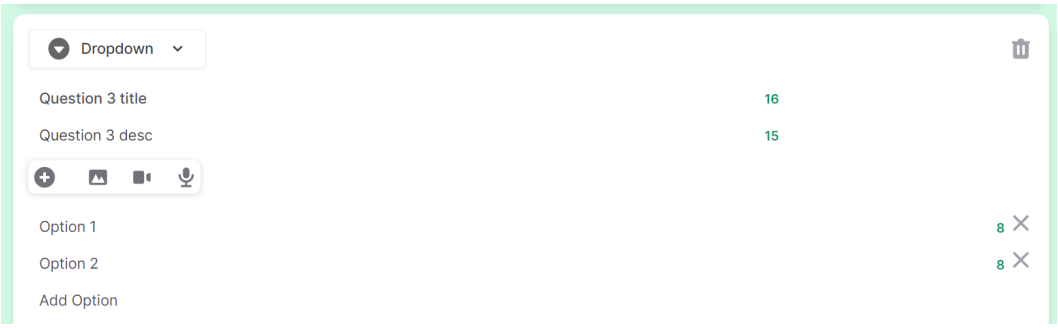

Figure 6.1: Question Creation Component Testing

In component testing many components have been tested but the most important components tested are the following. In the question component, the functionality for question

title, description is tested by providing stub data for the question title and description. We while creating the question the user also has functionality to upload the image, video, audio data. While uploading the image the component is tested by uploading image from the device or by uploading image via a url.

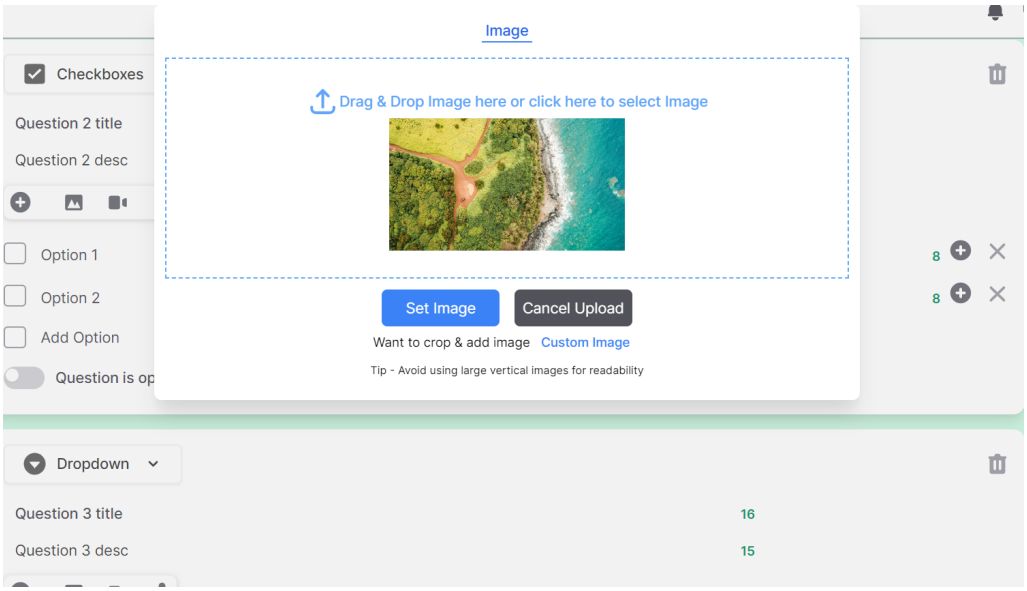

Figure 6.2: Testing Uploading Question Image

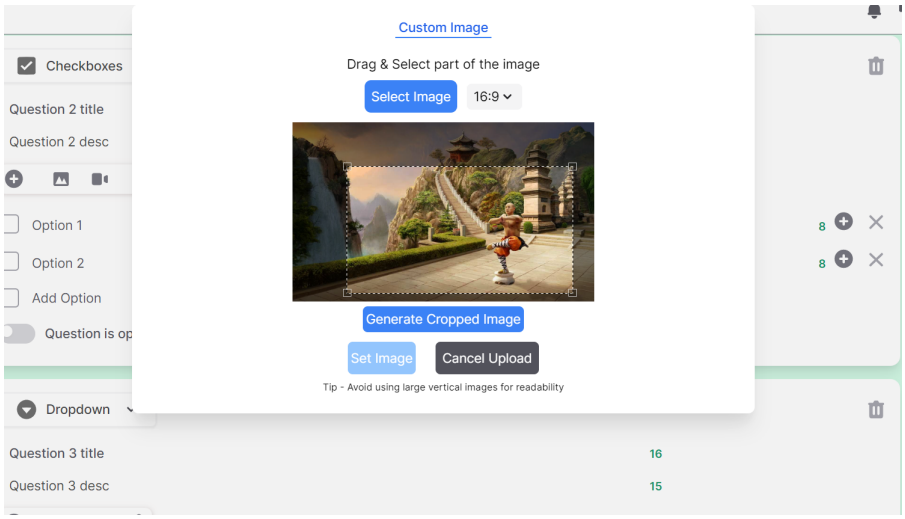

Figure 6.3: Testing Crop And Upload Image

The same is true for video and audio where the data can be uploaded from the device or via a url. The data uploaded should be properly visible in the upload model can be tested.

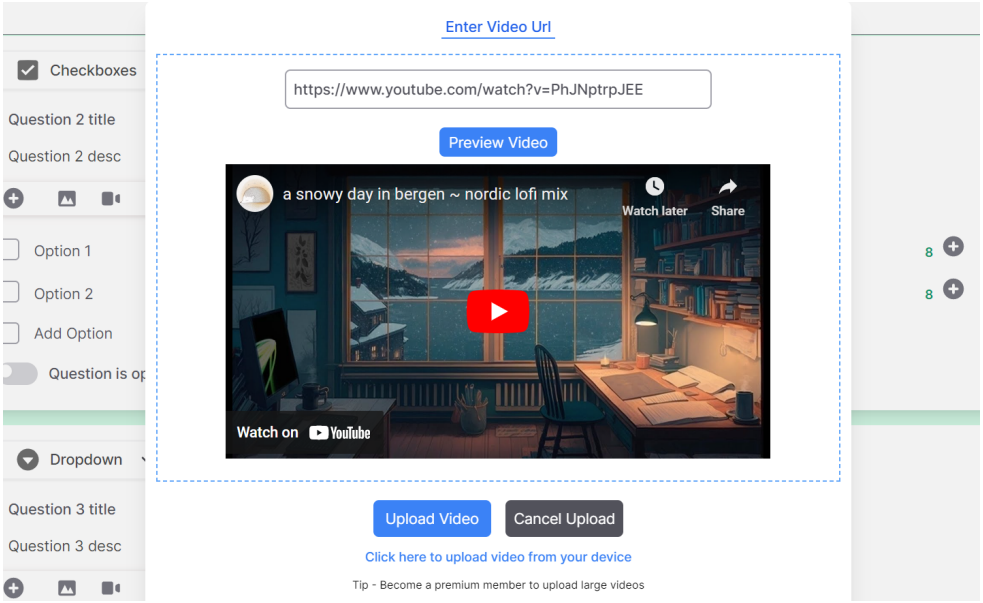

Figure 6.4: Question Video Addition Component Testing

6.1.2 Integration Testing

While modules and components can pass individually, quality engineers still need to ensure their functionalities as a group. Since a system's modules and components are commonly built separately by different developers, integration testing is critical to validate that they work together correctly. Modern software infrastructure often includes microservices that communicate with one another. These communications need to be included in integration testing and ensured that they operate properly.

Integration testing of multimedia data is done here. Once the multimedia data has been uploaded from the device or url to the app. It goes to the backend then it is stored in the cloud for fast retrieval purposes. Integration of frontend, backend and cloud is testing here and this test case will pass if data is visible in the form like shown below.

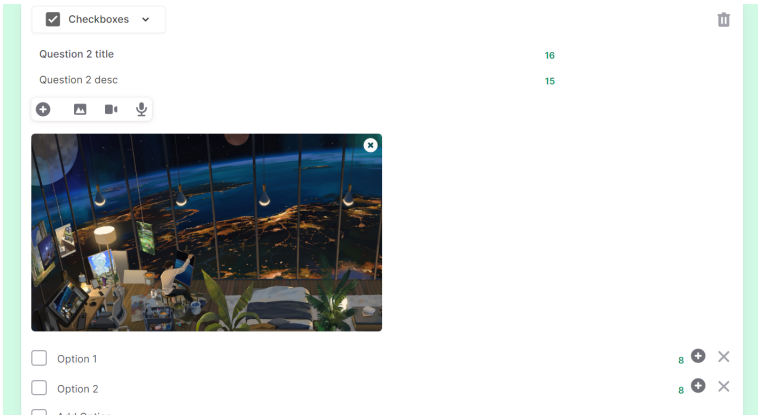

Figure 6.5: Question Image Addition Integration Testing

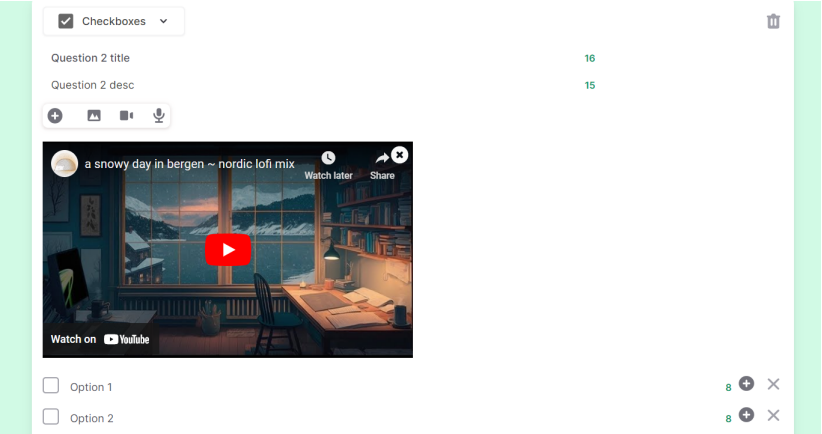

Figure 6.6: Question Video Addition Integration Testing

6.1.3 End To End Testing

As the name suggests, in this phase, the software is tested as a complete, integrated system to verify that all business and functional requirements are met. Hence it is also referred to as end-to-end testing and often occurs right before User Acceptance Testing. To yield correct validation, the test environment for system testing needs to be an accurate replication of the production environment. On top of that, it is performed in the white-box testing method, where testers have no involvement in the development of the system.

End to end testing is performed for the survey form creation page. Where the survey page is tested is for the following functionality.

1. New question creation :- Here the new question is created by adding the question title, question description which is optional. Along with addition of image, and other multimedia data by clicking on "Add Question".

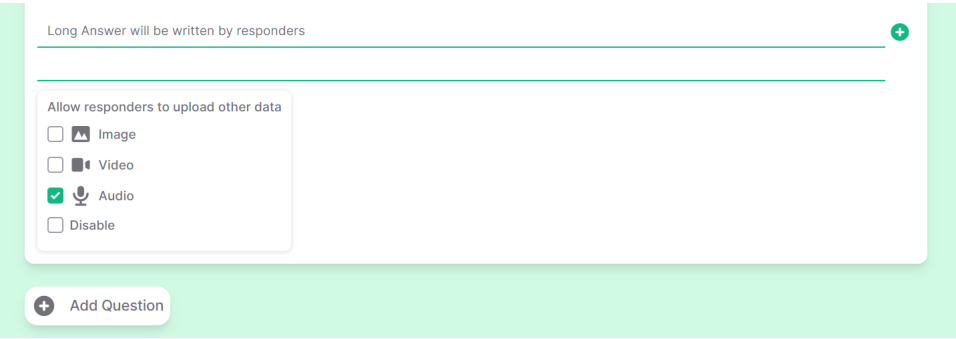

Figure 6.7: Testing Add Question

2. Question deletion :- The question is deleted by clicking on the trash icon. If only one question is present in the form then deletion won't be allowed. Otherwise the question will be deleted successfully resulting in deletion of it's text and multimedia data.

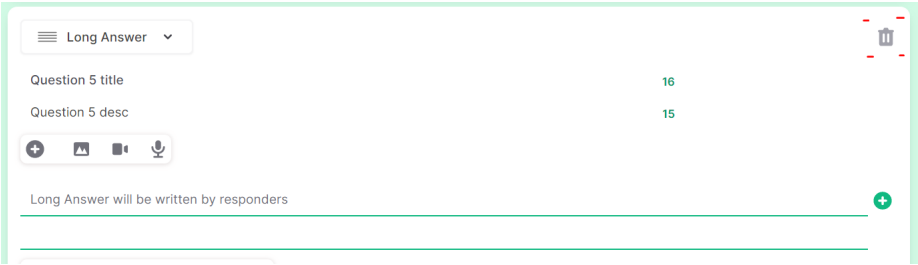

Figure 6.8: Testing Question Deletion

3. Question option creation :- Since the form supports choice based question types like "Multiple Choice", "Check-boxes", "Drop-down". The option needs to be created along with option text for these choice based question. This can be done by adding option as required

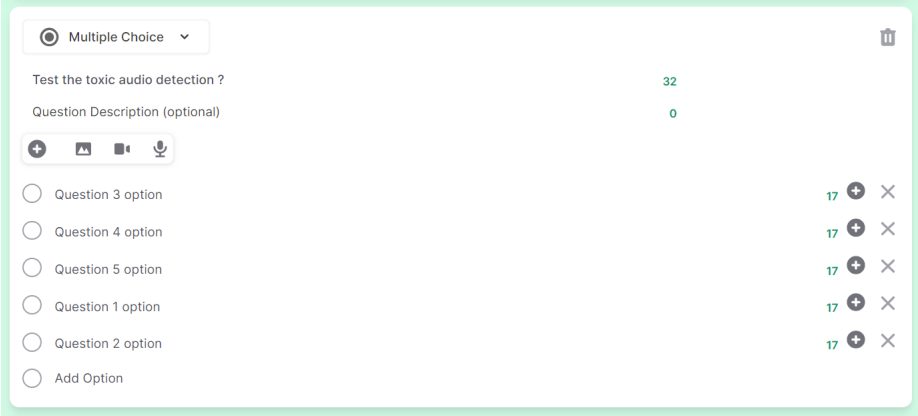

Figure 6.9: Testing Option Creation

4. Form publishing :- After adding all the questions that the form creator wants the form can be published by clicking on the "publish button" after which the survey will be visible to others and can be answered by others.

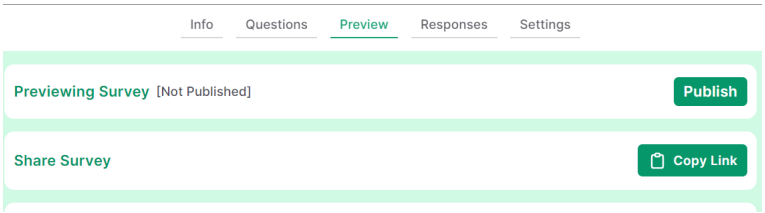

Figure 6.10: Testing Publish Form Functionality

5. Toxic content detection :- After publishing the survey the content sanitization request is sent to the backend for detecting and handling toxic content. More images on content sanitization are present in the Results section.

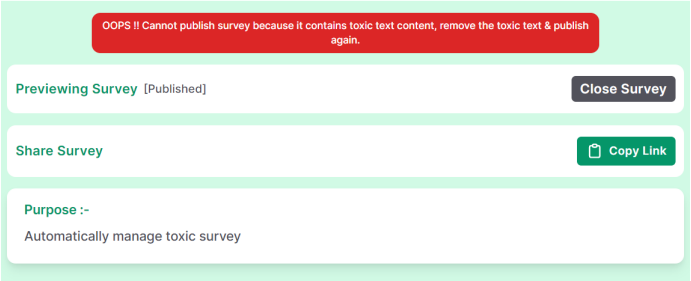

Figure 6.11: Testing toxic content sanitization

Sr. No.	Test Case	Expected Result	Actual Result	Result
1	Entering toxic text or swear word in the survey	Should show red alert for toxic text.	Red alert shown for toxic text	Pass
2	Entering normal text without swear word	Shouldn't show red alert for toxic text & submit successfully	Red alert for toxic text not shown and submitted successfully	Pass
3	Entering audio containing swear words	Should show sanitization message and swear words should be beeped.	Sanitization message shown and swear words beeped.	Pass
4	Entering audio not containing	Shouldn't show sanitization message & submit successfully	Sanitization message not shown and submitted successfully	Pass
5	Entering image containing nudity	Should blur the image and show sanitization message	Image blurred and sanitization message shown	Pass
6	Entering safe image	Shouldn't show sanitization message and submit successfully	Sanitization message not shown and submit successfully	Pass
7	Entering video containing nude frames	Should blur the frames & show sanitization message	Nude frames blurred and sanitization message shown	Pass
8	Entering safe video	Shouldn't blur the frames and submit successfully	Frames not blurred and submitted successfully	Pass

Figure 6.12: Functional Testing

Chapter 7

Results

The below shown diagram displays the underlying architecture for Machine Learning integrated with an existing client server application and it's effective use for sanitizing (handling) inappropriate text, images, audio and video.
The below figure shows the swear (toxic) word "fucking" which is present in the survey.

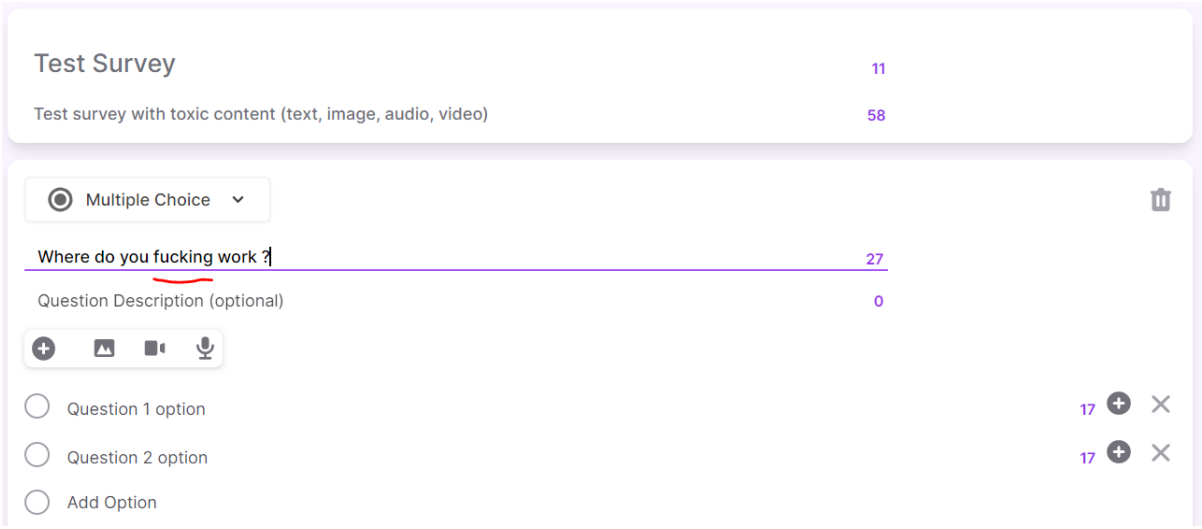

Figure 7.1: Toxic Text Present In The Survey

The below figure shows the publish button in order for the survey to be visible to the responders.

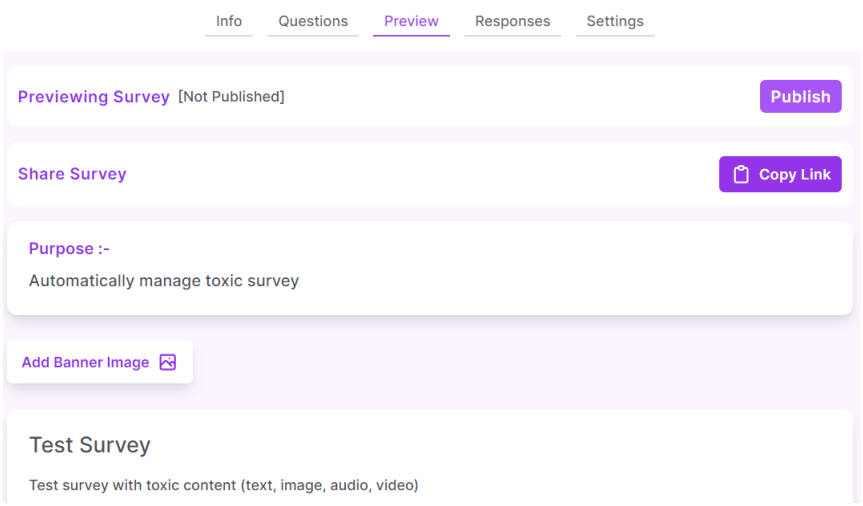

Figure 7.2: Publish Button

The below figure shows that on clicking the publish button the ML-OPS come into picture as the survey is added in the processing queue for processing the text, image, video and audio present in the survey.

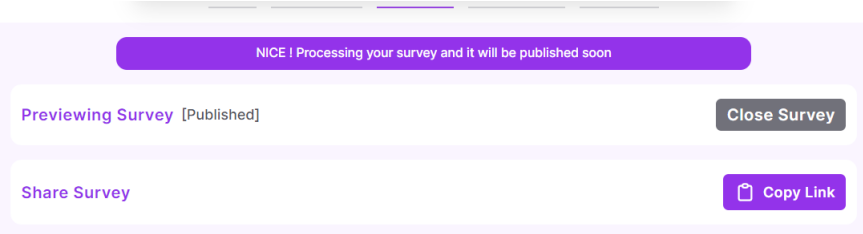

Figure 7.3: Processing The Survey

As shown in figure 6.2 the survey contains the swear word "fucking" the survey will not be published and a message will be shown in red.

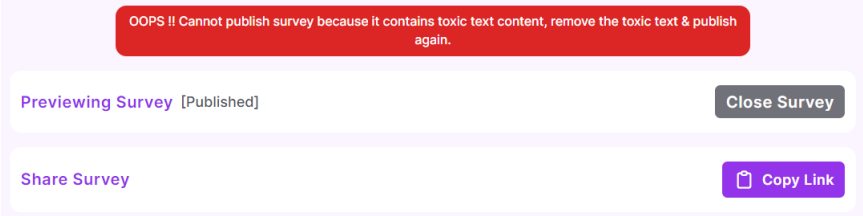

Figure 7.4: Sentiment Analysis To Analyze 6 Categories Of Toxic Text

The below image shows that audio file has been uploaded and the following text is spoken in the audio file. The transcript is as follows

"Top Cat! The most effectual Top Cat! Who's intellectual close friends get to call him T.C., providing it's with dignity. Top Cat! The indisputable leader of the gang. He's the boss, he's a pip, he's the championship. He's the most tip top, Top Cat. Bollocks ! He is the fucking champion ever to win a race 3 times consecutively. Mutley, you snickering, floppy eared hound. When courage is needed, you're never around. Those medals you wear on your moth-eaten chest should be there for bungling at which you are best. Bugger ! Stop that pigeon, stop that pigeon, stop that pigeon, stop that pigeon, stop that pigeon, stop that pigeon, stop that pigeon. How! Nab him, jab him, tab him, grab him, stop that pigeon now. Shut up you bitch don't talk with me."

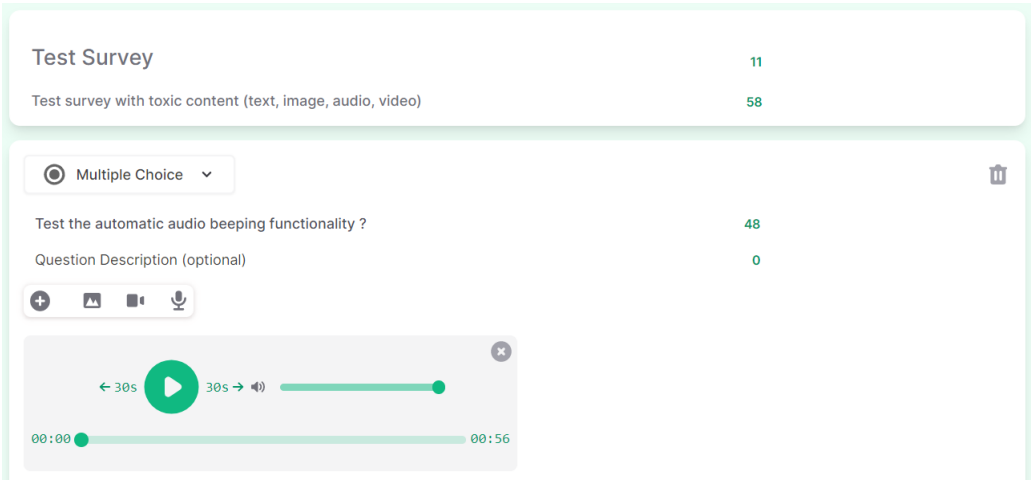

Figure 7.5: Audio File Uploaded With The Above Transcript

On publishing the survey again after uploading the audio file, the processing message will be shown above

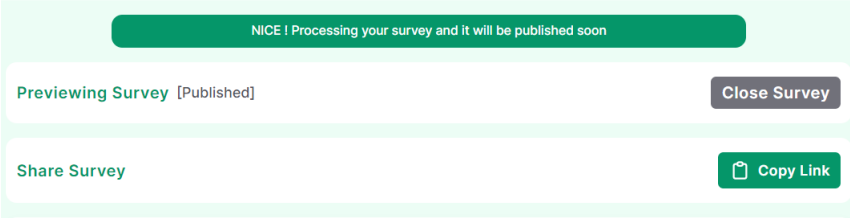

Figure 7.6: Survey Processing Message

As it was clearly evident from the transcript given above. There are swear words like "fucking", "bitch" in the audio. The audio will be processed audio where the swear words will be beeped.

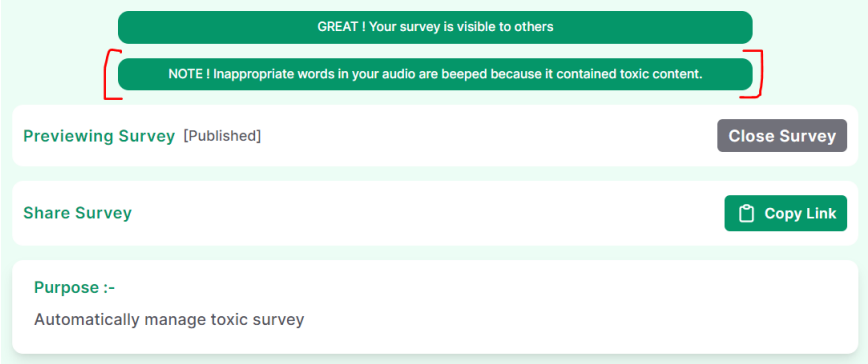

Figure 7.7: Swear Words In The Audio Beeped

If images containing any inappropriate views are uploaded then those images in the survey will be processed and if the images contain nudity, sexual intercourse scenes, or any exposed parts then the image will be blurred and the original image will be replaced with the new blurred image.

Note: As we cannot display the object in the above sample image fully but to understand the processing and mechanism of what type of images will be automatically detected and blurred we have used the above example.

Below image shows that the inappropriate image uploaded in the survey has been blurred.

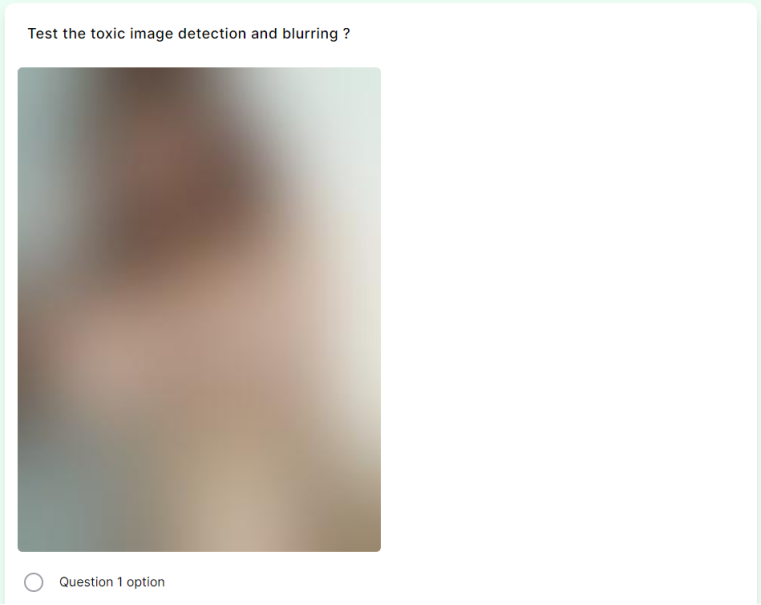

Figure 7.8: Processed Image (Automatically Blurred)

The same technique is used for video as well the video is processed frame by frame and if the that frame contains anything inappropriate then that frame will be automatically blurred. The result will be a video with few blurred frames and the original video will be replaced by this new video.

Chapter 8

Conclusions and Future Scope

Conclusion:

During the course of this project, we achieved text, image video and audio content sanitization and exclude inappropriate any multimedia data for our website environment by using logistic regression, Open CV, CNN Models, NLP, and Rabbit MQ. We also achieved request processing in bulk without dropping the requests in a Competing Consumer Pattern. This can be applied to large applications as well but the number of Consumers needs to be increased in accordance with the average number of requests. We hope that in the future this mechanism can also be used in various other platforms and technologies to reduce toxicity and to discard inappropriate content.Our scope is to properly analyze and sanitize all the multimedia contents so that the survey form owners or creators gets a good and appropriate data without having any inappropriate information provided to them and to keep our platform menace free so that any personnel will not be able to create havoc on our platform.

Future Scope:

In the future, we are planning to develop an ecosystem on our platform so that the functionalities implemented on our platform can be integrated with other platforms wherever required in real time. Many organizations will be able to benefit from it. Also, larger amount of data will be used for training and testing in future. There is also scope for increasing the accuracy of the model by scraping more data from the internet and using it for training our model. Also, the request processing time required by consumers can be decreased significantly by running consumers on machines with faster computer architectures.

Bibliography

[1] Clayton Santos, Eulanda M. dos Santos, Eduardo Souto,"NUDITY DETECTION BASED ON IMAGE ZONING", "https://sci-hub.se/10.1109/ISSPA.2012.6310454" The 11th International Conference on Information Sciences, Signal Processing and their Applications: Special Sessions,2020

[2] Rahat Shahriar Islam, Raisa Siddiqui, Dipta Roy,"Blurring of Inappropriate Scenes in a Video Using Image Processing".

[3] Shoji Kido, Yasusi Hirano, Noriaki Hashimoto"Detection and Classification of Lung Abnormalities by Use of Convolutional Neural Network (CNN) and Regions with CNN Features (R-CNN)"

[4] Bulbul Bamne, Neha Shrivastava, Lokesh Parashar and Upendra Singh,"Transfer learning-based Object Detection by using Convolutional Neural Networks ",Proceedings of the International Conference on Electronics and Sustainable Communication Systems (ICESC 2020) IEEE Xplore

[5] Midia Yousefi and Dimitra Emmanouilidou,"Audio-based Toxic Language Classification using Self-attentive Convolutional Neural Network" paper was published on June 2021."file:///C:/Users/prana/Downloads/IEEE_Conference_Template_Toxcicity_detection_Eusipco%20(1).pdf"

Project Achievement

Project titled **"Content Sanitization Using CV and NLP"** was presented at the inter-college university level project competition **'Exalt 2k23'"** by **"Parth Bhoir" "Pranav Mayekar"** and **"Anjali Singh".**. It gives us immense pleasure to convey that we have secured **"2nd Rank"** in that competition and look forward to further improve the project.