

# Infosys

---

## Responsible AI Office

### Infosys Responsible AI Toolkit – Safety tenet API usage Instructions

#### Contents

Introduction.....	2
Analyze.....	2
Censor .....	3
Image Analyze.....	4
Generate Image .....	5
Video Analyze .....	7
Nude Image Analyze.....	7
Nude Video Safety.....	8

# Introduction

AI models can sometimes generate harmful content, such as profanity, toxic language, explicit images, or sexually suggestive text. To ensure safe and responsible AI, it's crucial to implement measures that filter and prevent the generation of such content. This involves using advanced techniques to detect and mitigate harmful outputs, protecting users from exposure to inappropriate material, and maintaining a positive and inclusive online environment.

Once API swagger page is populated as per instructions given in the github repository Readme file, click on 'try it out' to use required endpoints. Details of endpoints associated with Safety tenet are outlined below.

## Analyze

**Endpoint** – /api/v1/safety/profanity/analyze

Using this API, we can check if the text contains any profane words or not and we can get the toxicity score for the same.

### Input :

Replace the 'inputText' with the prompt we want to check for any profane words.

POST /api/v1/safety/profanity/analyze Analyze

Parameters

No parameters

Request body required

```
{  "inputText": "You are a dummy",  "user": "string",  "lotNumber": "string"}
```

Execute

### Response :

Server response	
Code	Details
200	<p>Response body</p> <pre>{   "profanity": [     {       "profaneWord": "dummy",       "beginOffset": 9,       "endOffset": 14     }   ],   "profanityScoreList": [     {       "metricName": "toxicity",       "metricScore": 0.986     },     {       "metricName": "severe_toxicity",       "metricScore": 0     },     {       "metricName": "obscene",       "metricScore": 0.002     },     {       "metricName": "threat",       "metricScore": 0     },     {       "metricName": "insult",       "metricScore": 0.979     },     {       "metricName": "identity_attack",       "metricScore": 0.001     },     {       "metricName": "sexual_explicit",       "metricScore": 0     }   ] }</pre>

## Censor

**Endpoint** – /api/v1/safety/profanity/censor

Using this API, we can censor any profane words identified in the text.

### Input :

Replace the 'inputText' with the prompt we want to censor for any profane words.

POST

/api/v1/safety/profanity/censor

Censor

Parameters

No parameters

Request body required

```

{
  "inputText": "You are a dummy",
  "user": "string",
  "lotNumber": "string"
}

```

Execute

## Response :

Server response	
Code	Details
200	<div>Response body</div> <pre> {   "outputText": "You are a ****" } </pre>

## Image Analyze

**Endpoint** – /api/v1/safety/profanity/imageanalyze

Using this API, we can check if the image falls under any of the following labels – drawings, hentai, porn, neutral or sexy.

### Input :

Upload the image to be analyzed.

POST

/api/v1/safety/profanity/imageanalyze Imageanalyze

Parameters

No parameters

Request body required

portfolio

portfolio

☒ Send empty value

account

account

☒ Send empty value

image \* required

string(\$binary)

Choose File profane.jpg

Execute

## Response :

Server response	
Code	Details
200	<div>Response body</div> <pre>{   "analyze": {     "drawings": 0.002193321008235216,     "hentai": 0.004320410545915365,     "neutral": 0.9761048553324166,     "porn": 0.015960555523633951,     "sexy": 0.001420825021341443   },   "ORIGINAL": "1VB0RwKGG0AAANSuHEUGAAAKMAACUCIAAAABwC7CAACKB00IEQVR4Nk92Bw1axifiln7tBrytrVlbUzmrTxbOCdabB1LSbwICaEyDjauoj9CSTTD760IvezfoL9KIHwU5g25S2jR6G4lmsJBDYg0BIDY6K26urasqlzveU7Z YwM9d9H8SfzRQpWgSIA/n2ybtCtG6EeH+Pb78J0ce+3//+dWdQRP/FGAlt0m9+CP3pJwHhPaIQFEGNJOt9KfPAUKTm9vqf4M82y472N5Q4ep1KREAL1AASGp+eAE3IFzXzb1C9+Mw108B0QTEZVC1IFxAgIDghBwY1S30QcL230V+QdPj6 YysKXAT79tV2Fdz78TtY2P8FVQ0QAICoIASv0kch/2SE1CKAAqNT0KrkogBzj1MmC1nVEBVQH3VRVZ11qKulc+9wYHGM13T+h+d/SM/9zms/vxy/gkhQAACQFxxw16ftYmq1KFAADML31K/vYHhA1AAoqsVopJARTku2J0Kf1gKSK7MG0S FEATIZ7JwATYCS0LTf2Pysf+80BER8Np1+YU1NcJB4a1IABZ/1fcSV9pe7B8HQf8dxVQ8aCYG85P8tRtHwAGfUJUMC08TmkyEL1Th1FCSPcSDIV31102H30kVp0SutuaRD2u+Sheudaf+IntaEQ+u8ZRH679/S81ZPv97Vd+qVp/6x04B8dn2Kpa ICkba8G6c4Sk5GqC9Yb24deUmbWYJ33hFk2wC0U/0k11VIG3R5YAX55ZusAfiqV//8nPhaxo6pwy/7AS/dwQ5Vd45QUBAJQUEBBRQApTm2GA+N8pX9m6GCGAAAVUurJ0k1kUQVRE4RYT6rcim11se0g2h5ch555cgCAKSIQAARFfH5Kb/+TV675X8 8UPaNF1GHn2g1+4/8V0KBRACL5U1A1AUCBC5Xs/q7Fh3M/dQJ4K0BGAR2M525ArC8InLZ517De0Anrwi1q12KiffgYboc0AogABEGgKCUVCVua/4vhqhtn1xV31PbLBrpxrerg2UgEUSWqD77a8D9U1V2bu/hfBAAgJRFg5tq25AUz1LzumsutK8 uht1ZmF9FwVwIAAABW8Za1a1nT+d95UJ10M/V11C8mP66248C9FxFdu2wVQcBPvmeyno1p8sarthV9vUgagvntZ2ms1x0TVRQ1JUK11brooqr8MSXfmxh861CJ0U30M1K/7W61ACQmZnzUz17rqumw4p20m2599M1H1AVT 7//w98bx4790/ITZjCLB8XQgR2p2eugJpAUe11Q3yKtWwR100D28WUEJ37m9ah1/2eb20mH+smvG1SVB1R8hNkRj1Igyb1WkAgpWKCfHrCKe3jUdbQFT5Agp0Z/m011/d3VwvTAA552a2BrSmEYJH913bn7n3s92pvaM1+793dpl462 L3fCUTsA1h1ghaQkyqJ0hahSPH0XZrFudIvJk9e7UBA1St1zgeQpUUK8GKYToEYacRqgBuz2pcZmZjmfF8Zln/+tX/1M/pJue4B1CGfuv7A0Zd5cK5oPpT0d4FHLy8ScpwV1uzpfnDyzVIMEjo/vM6ZLmbdJcU1A01pQ1Yggv1tVn8 +7/VmB1a86Rc12V/wQ5REHL707558Tafu3CBCT16ydxPpJn10XkFEG2QMUH38qV9QUZAXG2/u4M1rv50110KKZgkq130VY3z/CP2+btKduZ5/On1oe7R5yf3JUXR5+4P3WwAB88P6jBkRZgKGCX/588P8bvp12VB+0Nvh379qz79+q1A5191UfP +enK8MSxM/76XUL0ePH378yftfM0ZLV2928WU1kosP/C3f/kp+FF7P11gbF8AgTAKS0LQhbyuk/52azfny8v5U9fFev011+vz07+8T/8Tx/97Ef701C9Z80bK4/PD91utgatWELUVEfHYCAFUBTFxhyFIV4GJ4FL71N0Q/39fPqERbaJg3mdu18sh7 72nF996V2Q04500B5fUDUSGwG0C1emG0K3/3Y87zeV1CV/5KfQf08JwH0MAEA1xagukotqb3t+9v/277XU5LO+Xp09HFqny+Nm01uv93PgeYfQhKwIAKQZysvhl1+LLZ/YQ9F0w7Z09mUZas19ZtQ3y/nJjzAMzbr2f10t154mbk5GZEFU VAmKycs3j3V12N2D1x+K1F0RcP16Goyangj1vum20240V1/R3m377/9Hf4AE8B51j1N1B+2n2me25Nq2AS2YGB+NDvY1fayfXk7xJ0WdW00B8BgnP8K6SfKdG35s+0T32/39wvVwK6GwWcYjht379yngGQ0WMLR859V74 3BwAElB/64U/pd1KXG134Q1k8oyQcSR230x0q21KwH+13Vd33/5VTV22bTtvd26r7/97W+TqgH1gvp13K20BQVEEVEHQK16bTryaxLm/AHt/3VR+7mPv751MBURFRT9F6n0LwH+1SPF34N/Um+/CHF/5P/uk/+0W+rv3vV26A005vEELCbhy ksjw7J3LYf35b+Ph2Y0oqJjw0K5KwUQVR1cgwyk/dwCV2H082bx5F2AJ0CSBK1BcA36ar9SPcas0hF0VfEPIITf0VtmykK9U180bdz/+2Q/h0VgtG+/K2PvHO/v7K8nc28w0Q0gXVqH6Uzhaiulqj1nnn/slvqr/ITpc181m7uH8RE RURf8NQCFktZk+FF7ZsmP5T/vfvyvuvvZYBE21CESXT+WMQLS4XZa3ue2tFHWVRFBAD1RWKpDInvU6bCPIsK0Hw0wV9ERjKAB101zASKGpSeseqdkICEqkkn31FUIU0dc/ep/BwcFh23eYOM1tJKABGGA1AaNgIFgP8daK101Jz+s3z 44U7v9L48Byt0aKCLdCpT49Jm4UHI+FL5MEHD77X9H+3d0eeqBEwckabD3puggp11Fe10K70xZV1Q03CLBd+AVULvZJRAEXV5K8Vf81QKEIT9/Ug00SZ5Vc15q21y1kH1t2zbNfY/PCJb/deup8Yq8tjpxUwBVZUwK0Ynpd727c802 ynb/C60H93541E2831Y4VJevTV9MLpWqJ7TE8HCU5GmSGQWAE1WocQULvAhrqU1K1LPB+Sf96V/IQdQVnrt4yFCA8EFJ7JTE8WgYd48G8S+S57LmH8F0JhJCOEupM001H2K8F8Sh0h0H11PvUwFdy 3y0v0694cd72W8KRTQdyA4a6gcQ1g1H1TmQC1TQC2CE7QREUB2U0V97ck11K15y1KZU0G1Q/1X0H1G5EPKX1ABUW8Ckq1J00Q0V21w1c1B4Ful570PpH19K1DVAf781TE8e8mgy20p8BENGSc2pNFcp039/8znZ/1...NO).....</pre>

## Generate Image

### Endpoint – /api/v1/safety/profanity/imageGenerate

Using this API, we can generate images based on the prompt and can check under which label(drawings, hentai, porn, neutral or sexy) they would fall.

### Input :

Enter the prompt to generate the image

POST

/api/v1/safety/profanity/imageGenerateImageanalyze

Parameters

No parameters

Request body required

portfolio

portfolio

☒ Send empty value

account

account

☒ Send empty value

prompt \* required

string

Generate a beach bikini image

Execute

**Response :**

[illegible]

# Video Analyze

**Endpoint** – /api/v1/safety/profanity/videosafety

Using this API, we can check under which labels the uploaded video belongs to – drawings, hentai, neutral, porn or sexy and can mask those profane objects in the video.

**Input :**

Upload the video file to be analyzed and masked

POST /api/v1/safety/profanity/videosafety Videoanalyze

Parameters

No parameters

Request body required

multipart/form-data

video \* required

Choose File

MicrosoftTeams-video.mp4

Execute

Clear

**Response :**

Code	Details
200	<div> <div>Response body</div> <div> <pre> {   "videoanalyze": {     "drawings": 0.0006344797438941896,     "hentai": 0.0001262754328683973,     "neutral": 0.9987052213869654,     "porn": 0.0004847407981287688,     "sexy": 0.00048024867282947525   },   "blurredVideo": "AAAAIG20eX8Pc29TAACAgIzb21pc28yYXZjJm1wDEAABSBm9vdGAAAGctdWkAAAAA... </pre> </div> </div>

# Nude Image Analyze

**Endpoint** – /api/v1/safety/profanity/nudanalyze

Using this API, we can check if the uploaded image contains any nudity and can blur the same.

**Input :**

Upload the image to be analyzed and blurred.

## Execute

[illegible]

**Input :**



