



Deepfakes Analysis Unit

Second Quarter Report 2025

July 1, 2025 - September 30, 2025

Index

- [About the DAU](#)
- [About the TIA](#)
- [How the DAU Works](#)
- [What Does the DAU Check?](#)
- [Motivation for This Report](#)
- [Scope of The Report](#)
- [Limitations](#)
- [Aggregate Statistics:](#)
 - [Type of Media Received](#)
 - [Domains Contained in URLs](#)
 - [Media Type Received By Week](#)
 - [Verification Status Assigned](#)
 - [Language of Media Shared](#)
 - [Themes in the Content Received](#)
 - [Media escalated by Fact-Checking Partners](#)
- [Appendix A: Verification Label Definitions](#)
- [Appendix B: Methodology](#)

What is the Deepfakes Analysis Unit (DAU)

- The Deepfakes Analysis Unit (DAU) is a collaborative task force set up by the Trusted Information Alliance (TIA), formerly known as the Misinformation Combat Alliance (MCA), to ensure timely response and research on the emerging crisis of AI generated video and audio.
- The primary public touchpoint is a WhatsApp tipline that allows the public to escalate purported synthetic audio and video to the unit for further investigation and action.
- The unit is helmed by an independent secretariat within the TIA that analyses content, escalates it to external forensic and detection experts, publishes assessment reports, and coordinates responses on the tipline as well as with fact-checkers.
- The DAU also has a separate escalation channel for fact-checkers in India. Earlier this year, it opened up a similar channel for IFCN-certified fact-checkers from across the globe.
- DAU also produces content for literacy around AI generated videos. DAU aims to be a trusted resource for the public, fact-checkers, and media, that detects and responds to misleading and harmful A.I.-generated audio and video.
- The seed funding to launch the DAU was provided by Meta.

Who are DAU's Fact-Checking Partners

BOOM



FACTLY



विश्वास News



the quint



newschecker

**TELUGU
POST**

Who are DAU's Forensic & Tech Partners



Detection and
Forensic Partners



Technical and
Research Partner



RIT

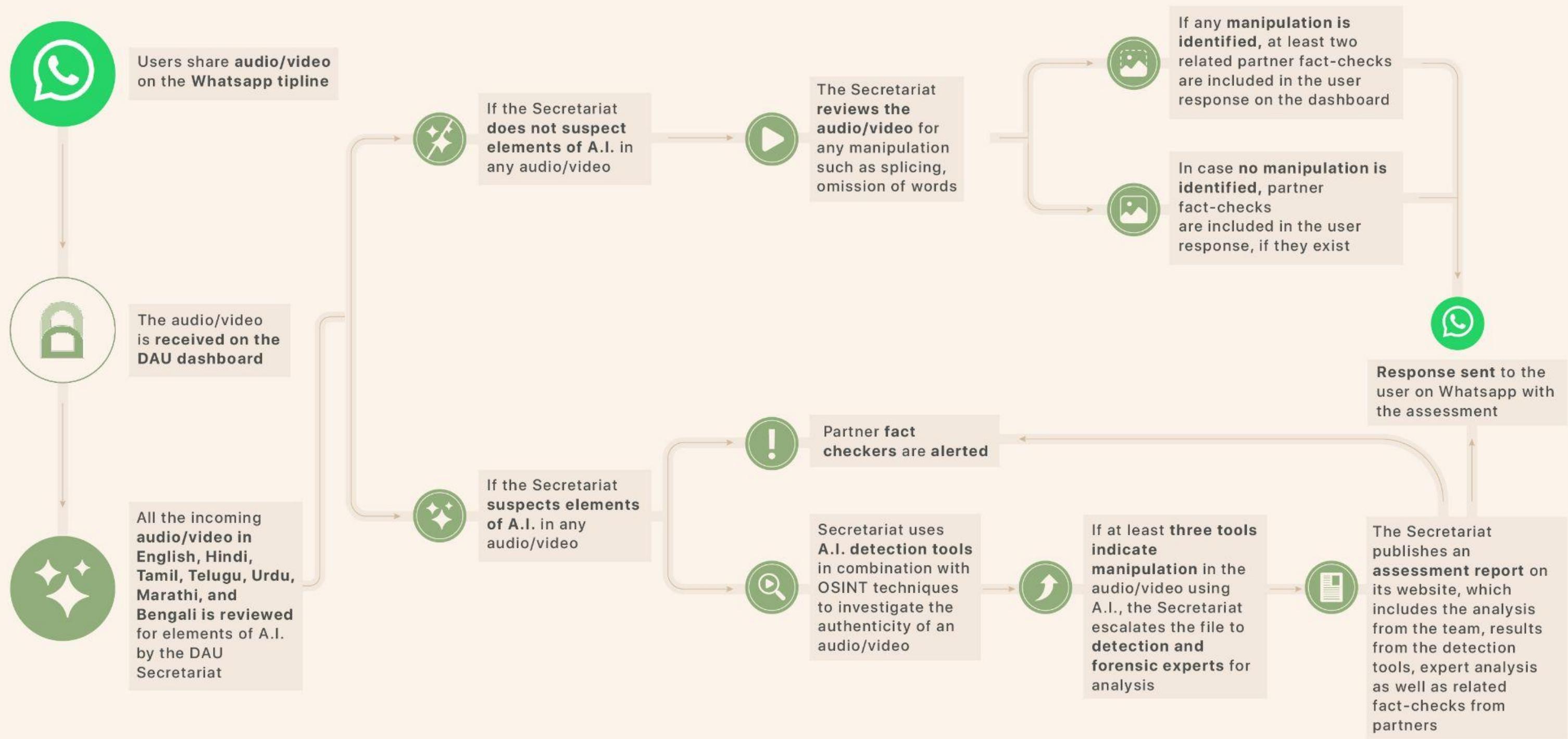


What is Trusted Information Alliance (TIA)

- The Trusted Information Alliance (TIA), formerly known as the Misinformation Combat Alliance (MCA), is a Section 8 not-for-profit corporation in India set up with the aim of collectively fighting misinformation and its impact.
- It is a cross-industry alliance bringing companies, organisations, institutions, industry associations and government entities together to combat misinformation and fake news and create an enlightened and informed society.
- The TIA at present has 11 members from fact-checking, media and civic tech organisations.

How the DAU Works

- Every message sent to the WhatsApp number that contains a URL, or a video or audio file is sent to a dashboard.
- The DAU secretariat looks at every media item on the dashboard. Images and text messages are filtered out and not sent to the dashboard. See the following page for a detailed description of the DAU's process.
- If the audio/video content fits within the DAU's remit (see next section), the secretariat conducts preliminary assessment using a manual review process as well as tool analysis. It checks for signs of AI manipulation as well as generation, such as imperfect lip-sync and oddities in facial features of a featured subject. It consults with experts to get further insights, if the initial investigation points to AI elements in the content.



What Does the DAU Check?

- At present the DAU only verifies audio and video content in Hindi, English, Tamil, Telugu, Urdu, Marathi and Bengali.
- Furthermore, content that is not in the public interest or is private in nature is not assessed and is considered beyond the scope of the DAU's focus.
- Pornographic or sexually explicit videos, however, go through a preliminary check by the secretariat to rule out the possibility of that video being Non-Consensual Sexual Imagery (NCSI).
- One of the three labels is assigned to an audio or video that falls outside the purview of DAU's focus:
 - Spam
 - Out of Scope
 - Unsupported Language

Motivation for this Report

- AI-generated content, including deepfakes, are an emerging challenge to a healthy online conversation.
- There are several speculations about the volume of AI-generated content online, its impact on the spread of misinformation and the ways in which the technology may be misused.
- This report is an attempt to add evidence to understand how AI-generated content is evolving and affecting online discourse.
- This is the second report of 2025, corresponding to the sixth quarter of the Deepfakes Analysis Unit's operation. To read more about how the DAU works, please visit [the DAU website](#) or read [reports on previous quarters](#).

Scope

- This report provides aggregate statistics regarding the content that the DAU saw on its dashboard during the sixth quarter of its operations, corresponding from 1 July, 2025 to 30 September, 2025. Previous quarters' report can be found on [the DAU website](#).
- The number of media items that the DAU sees on the dashboard is smaller than the number of messages that are sent to the WhatsApp number- text messages and images are filtered out and not analyzed by the DAU.
- This report provides aggregate statistics on:
 - the media type (audio, video, URL)
 - language of the content
 - the nature of manipulation, if manipulated (AI generated, manipulated, not manipulated, cheapfake, deepfake)
 - the broad themes in the content
- Engagement with content produced by the DAU, be it assessment reports or the media literacy videos, are not covered in this report.

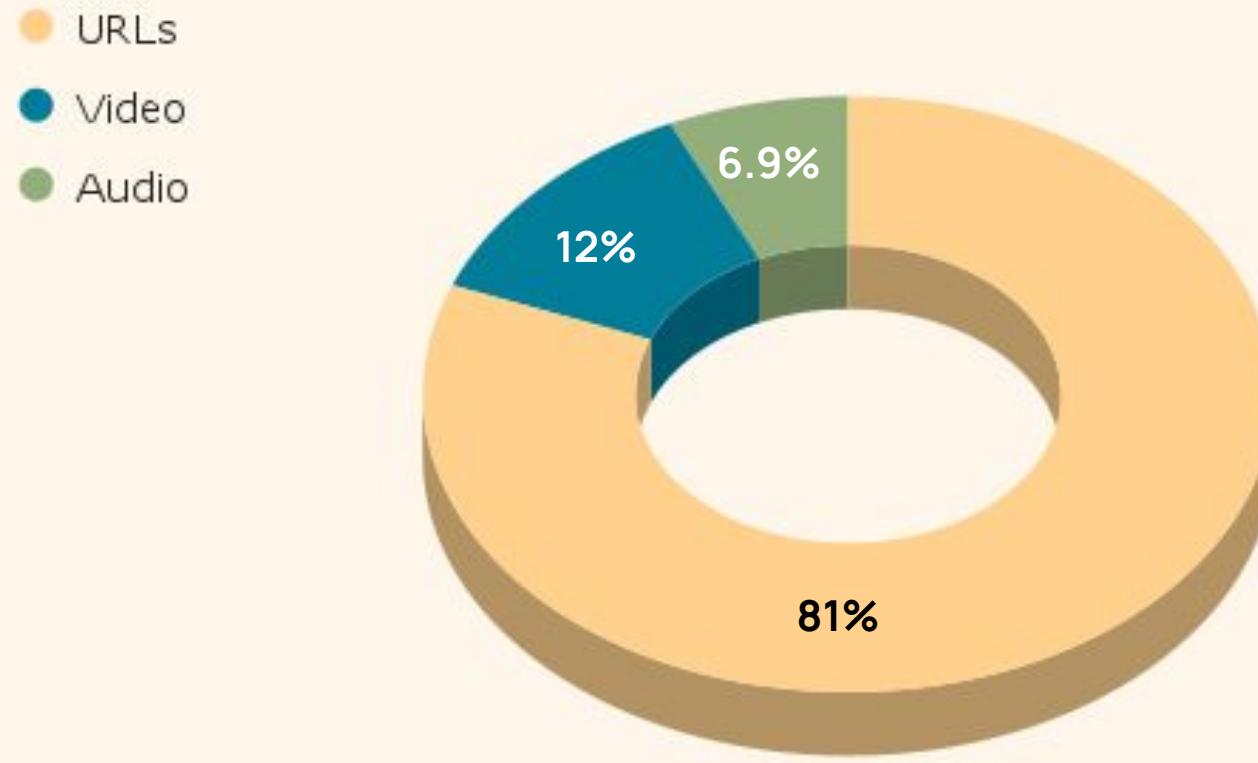
Limitations

- This report only shows data from one tipline focused specifically on deepfakes and content manipulated by A.I..
- There are several other tiplines run by fact checking groups that, while not dedicated to AI generated content, also receive it.
- The content hitting the DAU tipline should always be assumed to be a small percentage of the total AI generated content circulating online.
- The DAU database is not an exhaustive repository of all AI generated misinformation.
- Who uses the tipline, and how much the tipline is used depends on a number of factors such as ad campaigns to popularize the tipline, and the events at a time.
- For example, during a topical event where a lot of misinformation is produced, the tipline may be used more.

Aggregate Statistics

Please see [Appendix B](#) for Notes on Methodology

Type of Media Received

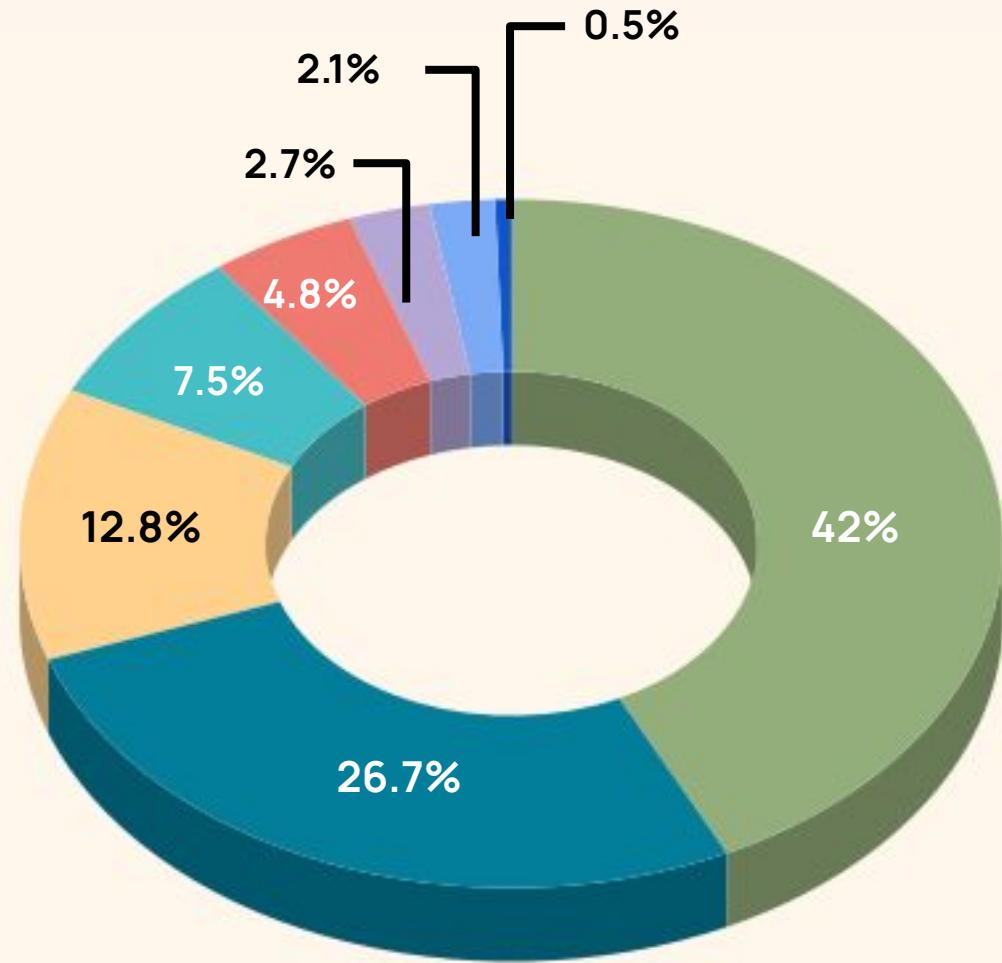


- Over 80% of the content was shared as URLs, indicating that users mostly submitted links.
- Direct uploads of video and audio made up 12% and 6.9% respectively, showing a preference for link-sharing over file submissions.

Total number of media items received: 189 (including spam)

Verification Status Assigned

- spam
- not_ai_generated
- out_of_scope
- not_manipulated
- manipulated
- ai_generated
- cheapfake
- unsupported_language

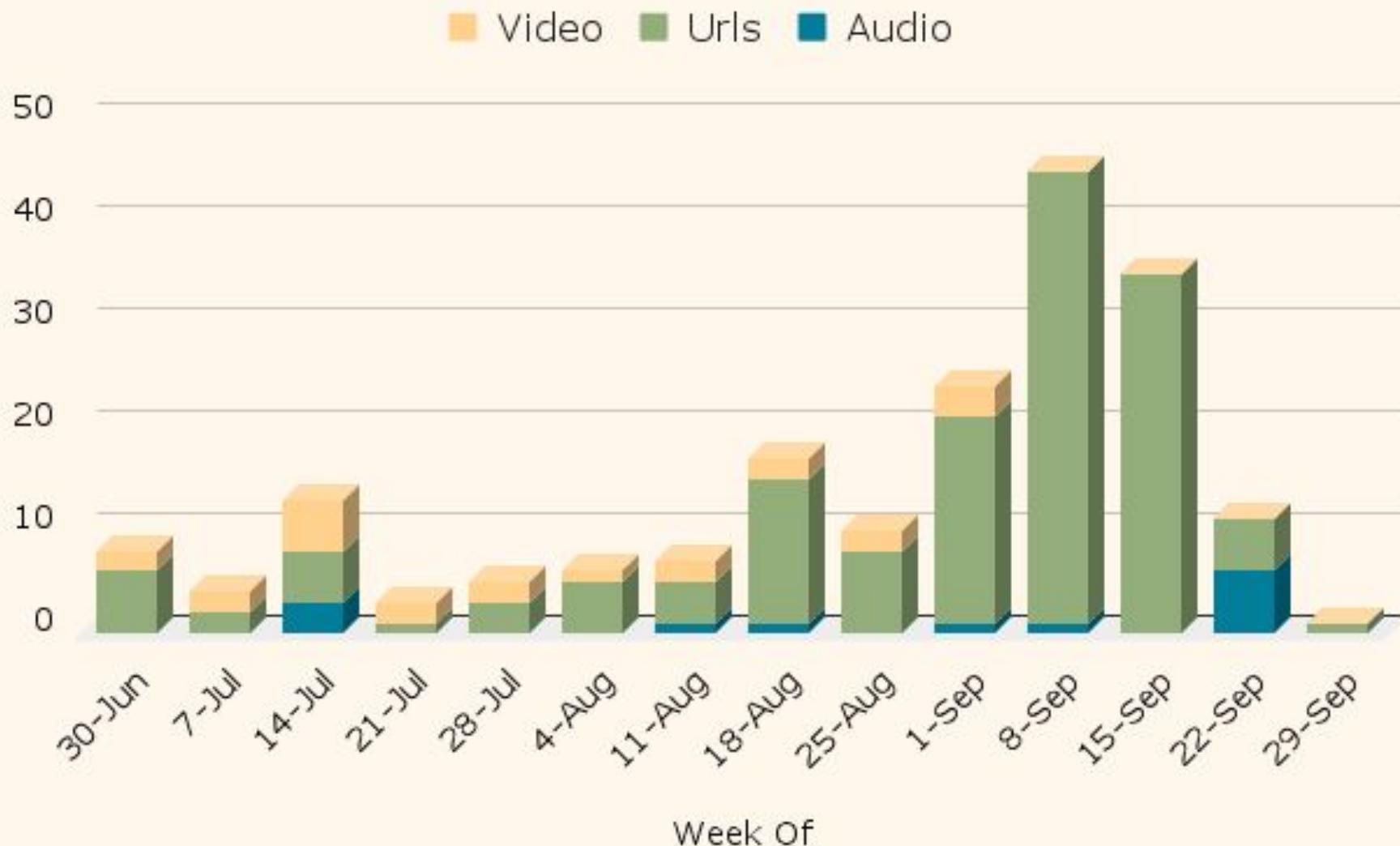


42% of the media items that hit the tipline were marked as Spam. Of the remaining:

- 50 items were labeled as Not AI-generated
- 5 items were labeled as AI-generated
- 9 items were labeled as manipulated
- No media item was labeled as Deepfake*

Media items labelled as Spam: 80
Media items assigned a label other than Spam: 107
Media items with a non-functional Url: 02 (Check [Appendix B](#))

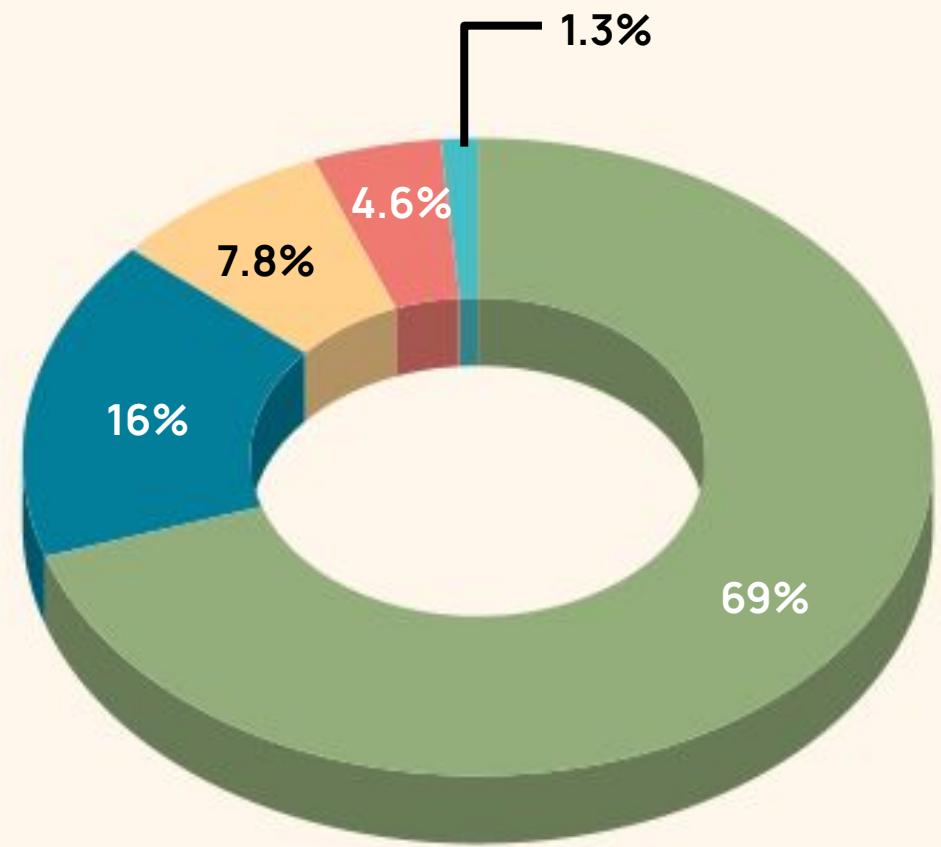
Media Type Received by Week



The highest number of URLs were recorded during the week of September 8: 44 URLs—more than double the weekly average.

Domains Contained in URLs

- Facebook
- Instagram
- Other
- YouTube
- Twitter

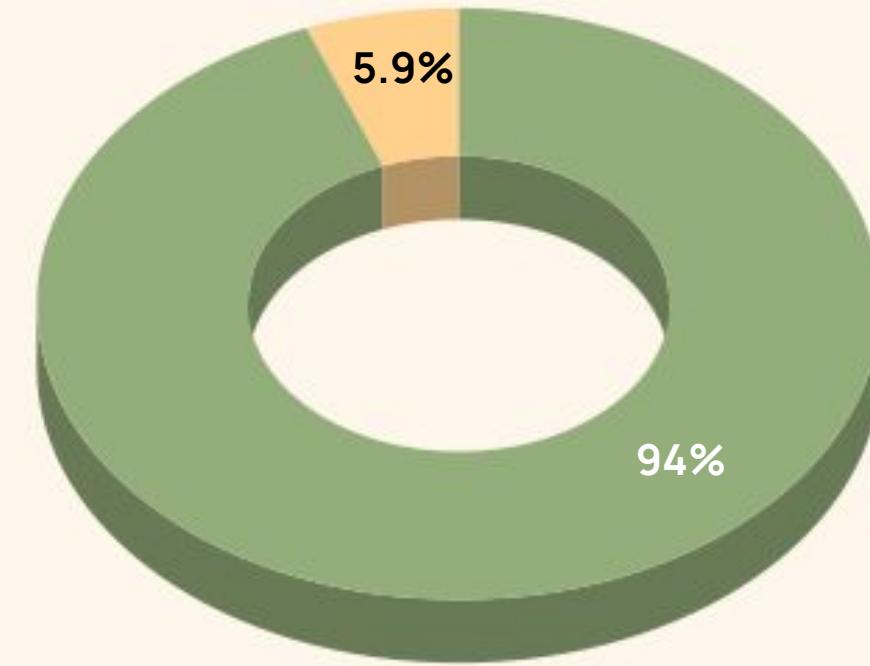


- Nearly 70% of all URLs came from Facebook, making it the most common source by a wide margin. It could be because of ad campaigns run by the DAU on Meta.
- Platforms like Instagram, YouTube, and X (formerly Twitter) contributed to 21% of all URLs.
- The “Other” category accounted for around 8% URLs, mostly spammy or low-quality domains such as 'm.par' 'hi.hello' 'goo.gl' etc.

Total Number of URLs: 153 (including spam)

Language of Media Shared*

- Hindi
- English



- The DAU added language tags (for seven supported languages) manually in media items that were not marked as spam.
- Over 90% of the media items (96) were in Hindi.
- 6 media items were in English.
- DAU didn't receive any media in Tamil, Telugu, Urdu, Marathi and Bengali in Q2 2025.

*These numbers do not reflect the language in which the user chose to interact with the tipline

Number of Media Items in DAU purview with a Language Label: 102

July 1, 2025 to September 30, 2025

Themes in the Content Received

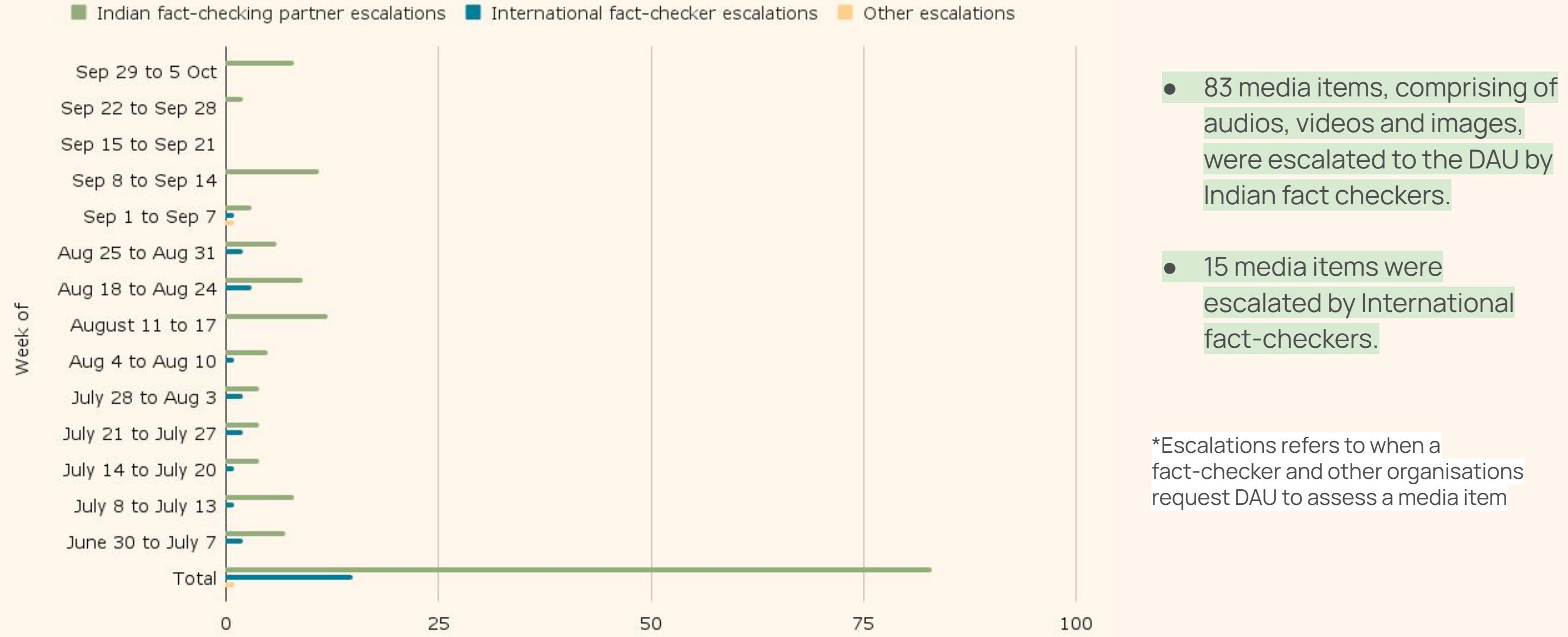
- Most of the videos debunked by the DAU during this quarter were A.I.-manipulated videos that surfaced in the wake of "Operation Sindoor".
- The individuals apparently featured in these videos were senior Indian defence officials such as chief of the army, navy, and air force as well as top Indian ministers, including India's Foreign Minister Dr. S. Jaishankar.
- The narratives in most of these videos peddled misinformation about India having lost combat equipment or defence personnel to Pakistan.
- Our expert partners from [RIT's DeFake Project](#) pointed to generation techniques that were likely used in some of these videos such as image-to-video generation, which involved the use of screenshots from a source video to generate video clips that were similar to the source video. In some cases that resulted in body movements looking different between the source video and the manipulated version.
- In addition to more obvious manipulations on the mouth and face of the apparent subjects in these videos, some common manipulation techniques included tampering of the insignia and name tags on the uniforms of the defence officials.

Continued to next page

Themes in the Content Received

- Another running theme in these videos was the use of distorted logos of media houses. The logos looked similar but were not identical, the discrepancies included different font sizes or typefaces.
- Most videos used voice clones or synthetic voices that sounded somewhat similar to the people apparently featured in the videos. In some cases some words that were part of the original audio track were used in the synthetic audio. We also saw a few examples of audio splicing, which involved stitching together parts of the original audio with the synthetic audio.
- Financial scams were only a small percentage of the videos that the DAU debunked during this quarter. The narrative in these scam videos focussed on fraudulent income-generating platforms supposedly powered by A.I.; and in some cases the videos claimed that the Indian government or senior ministers were backing these get-rich-quick schemes.
- The videos were similar to the financial scam videos previously debunked by the DAU in terms of the content as well as packaging, making them appear like public announcements or interviews for news segments. All of them involved the use of synthetic audio tracks with original footage sometimes of politicians, tech leaders or news anchors; the manipulations made it appear that they were promoting these schemes and urging people to invest.
- India's Finance Minister Nirmala Sitharaman and Road Transport and Highways Minister Nitin Gadkari were the politicians linked to such scam content. Television journalist Rajdeep Sardesai and Sundar Pichai, chief executive of Google and Alphabet, were among the non-political public figures purported to be endorsing these scams.

Media escalated to the DAU by fact-checking partners and international fact-checkers



Assessment Reports

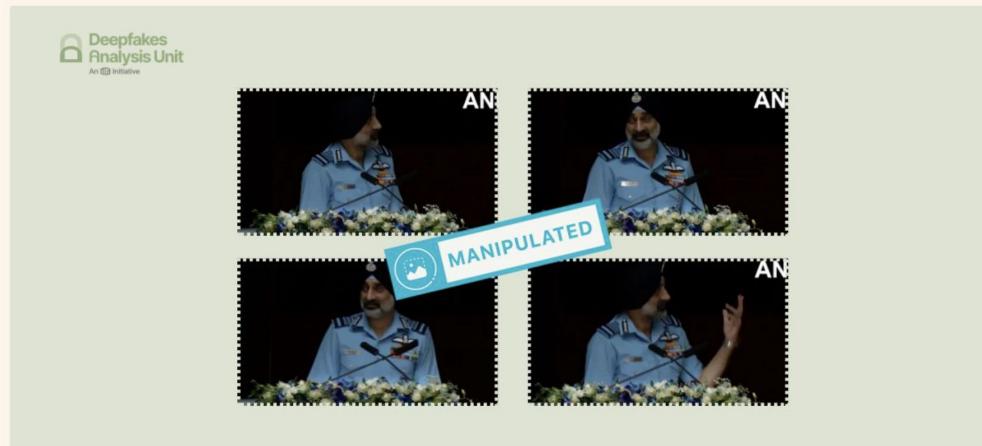
The DAU published 10 assessment reports in Q2 of 2025.

- 2 reports were based on media received via the DAU Tipline
- 5 reports were based on media escalated by a fact-checking partner
- 3 reports were based on media identified through social media monitoring

Video of India's Air Force Chief Amar Preet Singh Confirming Loss of Six Jets, One Drone to Pakistan Is Fake

August 28, 2025

Manipulated Media/Altered Media



Screengrabs of the video analysed by the DAU

Video of Sundar Pichai Promoting an Income Generating Platform For Indians Is Fake

July 7, 2025

Manipulated Media/Altered Media



Screengrabs of the video analysed by the DAU

Video of Deputy Army Chief Rahul Singh Confirming Loss of S-400 Missile Systems to Pakistan Is Fake

July 18, 2025

Manipulated Media/Altered Media



Screengrabs of the video analysed by the DAU

Feedback and Questions

Email:contactdau@mcaindia.in

Appendix A

DAU Labels and Definitions

DAU Labels and Definitions

For content that falls in the purview of the DAU, the secretariat assigns the media items one of the following labels:

- AI Generated
- Deepfake
- Manipulated
- Not Manipulated
- Not AI Generated
- Cheapfake

The secretariat produces an assessment report only for media items that may be detected to be manipulated using some element of AI. These could include media items labeled as deepfake, cheapfake, manipulated or AI Generated. You can read DAU Assessment Reports on the [DAU Website](#).

AI Generated

A video or audio, created using AI, that depicts an event, a person, or an interaction between people that never actually occurred or that alters the reality, *which may or may not mislead people*, is labeled as AI generated. If content has been produced with the consent of the person depicted in the video, or if the content is produced for humorous or creative purposes, it is labeled as AI generated.

For example, an AI avatar of a politician from Rajasthan depicting him speaking in Tamil was labeled as AI generated.



Example of a Video Labeled as AI Generated

Deepfake

AI generated content that includes fictional creations about a living or dead person, or a non-existent person, intended to deceive or cause harm is labeled as Deepfake. Common examples of deepfakes include swapping the face of a person (dead or alive) or cloning their audio to an extent where the generated content shows them doing something they didn't do or saying something they didn't say.



Every deepfake is AI-generated but every AI generated piece of media item is not a deepfake. AI generated content that shows nudity is labeled as a deepfake since the creator may or may not have sought consent from the subject(s) featured.

[Example of Videos Labeled as Deepfake](#)

Manipulated

Video or audio modified using simple editing software or AI technology to create a false or potentially misleading narrative, are labeled Manipulated. This includes cases where the original audio track from a video may have been replaced with AI-generated audio.

If the audio or video can be traced to an original source, the content is labeled as Manipulated, not AI Generated. In some cases, media artefacts such as the lips or mouth movement is blurred. This is seen as an attempt to synchronize the original visuals with the synthetic audio.



[Example of a Video Labeled as Manipulated](#)

Not Manipulated

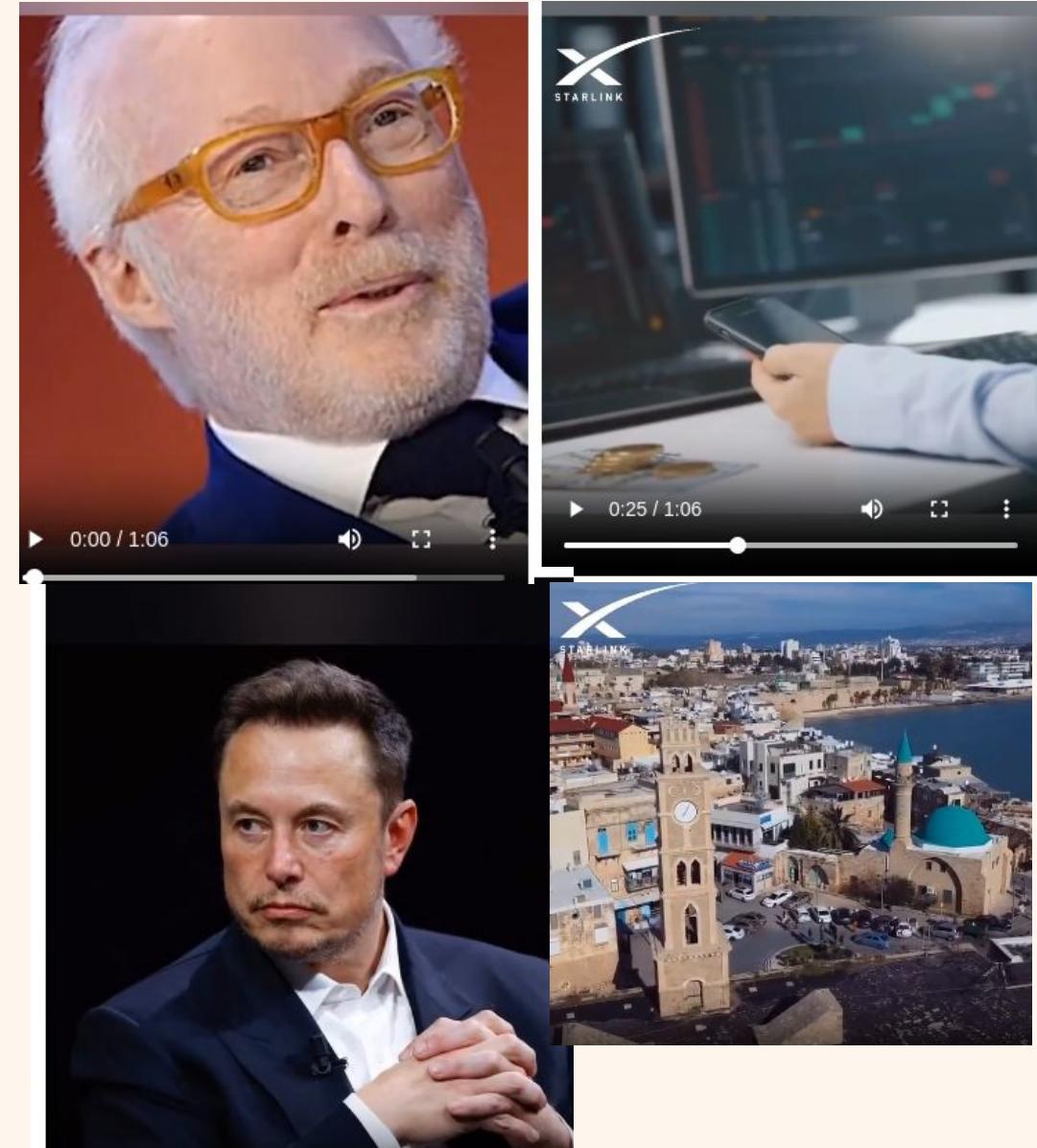
A video or audio which has not been tampered with and is identical to the source audio or video is labeled as Not Manipulated.

Not AI Generated

A video or audio, that is misleading, or has the potential to mislead but is not digitally altered is labeled as Not AI Generated. The DAU addresses these media items by surfacing fact checks, if any, from partners about the media item.

Cheapfake

A Cheapfake is a subtype of manipulated content where the quality of production is poor. This could include manipulation techniques such as slowing down, speeding up, reassembling, re-contextualising, and editing footage to create a false or a potentially false narrative; or satire or parody. The examples of content labeled as “cheapfake” by the DAU have included cases where the audio has been synthetic but the visual of a subject has been akin to a cutout or tucked in a corner of video with multiple other elements such as fleeting graphics or other moving visual elements.



Stills from a Video Labeled Cheapfake

Appendix B

Methodology Notes

Notes on Methodology

- The analysis is conducted on media items as received. Thus, the data may contain duplicates i.e. the same audio/video might have been counted more than once.
- Of all the content that is sent to the tipline, audio, video and URLs are filtered and sent to the dashboard for perusal by the DAU secretariat. While audio and video is detected through the file extension, URLs are detected by the presence of the terms: 'http' and 'www'. URLs that do not contain these terms or extensions of it ('https'), or are of an unsupported media type, do not hit the DAU dashboard.
- Of the media items that hit the dashboard, the following media items are considered out of purview:
 - Spam
 - Out of Scope: Content that is not in the public interest or is private in nature
 - Unsupported Language: Content in languages other than English, Hindi, Tamil, Telugu, Urdu and Bengali.

Notes on Methodology

- The secretariat attempted to assign all media items in its purview with a verification status, and a language tag. Where the URL was broken, that is the URL did not lead to a functioning website, the secretariat could not provide a verification status or a language tag. In Q4, DAU marked 36 items as broken URLs. Similarly, where the media did not have speech or text, the secretariat could not add a language specific tag.
- Users who sent spam content did not receive a response from the secretariat.
- When a media item is tagged on the dashboard with a topical tag, it is tagged with multiple tags. These tags may overlap with each other. For example a media item tagged with 'Amitabh Bachchan' could also be tagged with 'Celebrity'.