|  |  |
| --- | --- |
|  |  |

Text

Description automatically generated ISO/IEC JTC 1/SC 29/WG 03 N1235

**ISO/IEC JTC 1/SC 29/WG 03**

**MPEG Systems   
Convenorship: KATS (Korea, Republic of)**

**Document type:** Output Document

**Title: Exploration on indicating AI generated/altered content using the MPEG Systems technologies**

**Status:** Approved

**Date of document:** 2024-4-26

**Source:** ISO/IEC JTC 1/SC 29/WG 03

**No. of pages:** 8 (with cover page)

**Email of Convenor:** young.L@samsung.com

**Committee URL:** <https://isotc.iso.org/livelink/livelink/open/jtc1sc29wg3>

**INTERNATIONAL ORGANISATION FOR STANDARDISATION**

**ORGANISATION INTERNATIONALE DE NORMALISATION**

**ISO/IEC JTC 1/SC 29/WG 03 MPEG SYSTEMS**

**ISO/IEC JTC 1/SC 29/WG 03 N** **1235**

**Apr 2024, Rennes, FR**

|  |  |
| --- | --- |
| **Source** | **MPEG Systems working group** |
| **Status** | **Output** |
| **Title** | **Exploration on indicating AI generated/altered content using the MPEG Systems technologies** |
| **Serial Number** | **23845** |

INTERNATIONAL STANDARD© ISO/IEC 2015 – All rights reservedISO/IEC 15444-12:2015(E) 63Part 12: ISO base media file formatInformation technology — JPEG 2000 image coding systemTechnologies de l'information — Codage des objets audiovisuels — Partie 12: Format ISO de base pour les fichiers médiasInformation technology — JPEG 2000 image coding system — Part 12: ISO base media file formatE2015-02-20(60) PublicationISO/IECISO/IEC J   International Standard 2015ISO/IEC 15444‑ISO/IEC 15444‑12ISO/IEC 15444-12  Coding of audio, picture, multimedia and hypermedia informationInformation technology 291 2見出し 2見出し 1    02 STD Version 2.1c260   4C:\Users\shinji\_w\AppData\Roaming\Microsoft\Templates\STD\w15177\_14496\_5th.-restyle-R1.doc

# Introduction

This document defines the scope of a WG03 exploration of indicating AI-generated/altered content standards in the MPEG Systems standards.

The misuse of AI-generated or altered content is getting a lot of attention lately due to various reasons, including several incidents of deceiving the public by publishing manipulated media content on social media platforms that falsely appears to be authentic, including the audio and video of people appearing to say or do things they did not say or do, without their consent (‘deep fake’). Such misuse of AI has raised significant concerns among the public, media, and governments, and various legislative activities are being conducted to address the issue.

While defining the scope and definition of such AI generation or alternation, is not in the scope of WG03, the media system standard developed by WG03 can provide the means for marking if a piece of media content is generated by AI or altered by AI. Providing a technical solution, i.e. a clear placeholder for signaling whether the content has been AI-generated or altered, and perhaps the “marking information”, makes the standard useful for carrying information that would be required by law in different jurisdictions.

# Background and justification

The document JVET-AG0045-v1 provides an overview of the recent legislative activities in its Annex A. In summary, legislative works have been started in several jurisdictions that attempt to provide requirements for marking content AI-generated or altered content. The documented activities in that annex are:

1. The EU AI Act article 52 requires the disclosure of information regarding the AI-generated/manipulated content that may deceive the audience in an appropriate, timely, clear, and visible manner: “Disclosure shall mean labeling the content in a way that informs that the content is inauthentic and that is clearly visible for the recipient of that content.” and “To label the content, users shall take into account the generally acknowledged state of the art and relevant harmonized standards and specifications.”
2. The US Senate proposal (Schatz/Kennedy AI Labeling Act) requires the AI-generated content shall include a clear mark that the content is AI-generated. Furthermore, it requires “the output’s metadata information shall include an identification of the content as being AI-generated content, the identity of the tool used to create the content, and the date and time the content was created.” It evens goes further with “the disclosure shall, to the extent technically feasible, be permanent or unable to be easily removed by subsequent users.”
3. The National Information Security Standardization Technical Committee (“TC260”) released the final version of its “Tagging Standard” on August 25, 2023. The standard requires that for AI-generated content stored in the files, there should be added metadata for identification. This metadata should be in the format: AIGC: {“ServiceProvider”: value1, “Time”: value2, “ContentID”: value3} with specific guidelines on the length and format of each value (<https://www.insideprivacy.com/artificial-intelligence/labeling-of-ai-generated-content-new-guidelines-released-in-china/> ).

# Scope of this exploration

This exploration will explore how to mark media content with metadata that indicates the AI-generated/altered portions including the following aspects:

1. Collect and study the use cases and applications in which the marking of AI-generated or altered content is relevant.
2. Define the technical requirements of the above use cases. The requirements would cover the signaling and/or carriage of the information.
3. Study and catalog the MPEG Systems standards that are impacted by the above use cases and the relevant existing MPEG System technologies that can provide a solution.

# Use cases

# Considering the above legislative activities, we propose the following use cases.

# FF- Enhanced audio/video

|  |
| --- |
| **Name:** |
| Improved audio of ISOBMFF File |
| **Description:** |
| User Ada likes WWII history and watching news clips generated during that era from various websites. Since the files generated in that area often do not have good audio-visual quality, some service providers use AI technologies to enhance the audio tracks, especially when people are talking in outdoor environments.  Ada selects a video file and watches the content. The file contains an audio track that is enhanced using AI tools for better recognition of voices and clarity of words. The player whenever the audio is enhanced, displays an overlay “enhanced quality of speech with AI”. Ada can click on the player user interface and listen to the original audio tracks if she chooses.  **One or more of the following variations:**   1. Portions of the audio of the speaker are only enhanced. 2. Portions of the audio of the background are suppressed. 3. Portions of the audio of the background are boosted. 4. Additional closed shots are added to the original video. 5. The original file has only audio. An AI-generated video is created for the audio using archived videos of other events repurposed and altered to this event. |
| **Categorization** |
| Marking the alteration for enhancement |
| **Preconditions** |
|  |
| **Potential Requirements** |
| * Marking of the AI-generated/altered content in each ISOBMFF-enhanced audio track   + Information regarding the tools used, the time interval, the reference to the track with original audio content, the date, and the party responsible for creation and/or alteration.   + Easily findable and detectable information.   + Allows referencing one or more regulation schemes and/or regional laws.   + Support of multiple languages.   + Support of custom fields based on the applicable regulations. |
| **Feasibility** |
|  |

# FF - Deep fake content

|  |
| --- |
| **Name:** |
| Deep fake content in the ISOBMFF file |
| **Description:** |
| User Adam receives an email with a link to a file, saying interesting new clip. When Adam clicks on the link, an ISOBMFF file is downloaded to his device and a multimedia player is launched to play the content for Adam. The file contains a video track and audio track of 10 mins, showing a government official at a news conference. The content is played. The video and audio tracks are altered from min 5 to min 7:30, fabricating ‘deep fake’ comments that were never made by the official. The player detects the AI-generated/altered marking in the file, indicating that the samples in both audio and video tracks, from min 5 to min 7:30 are not authentic and are generated by AI. The multimedia player overlays the “AI-altered content” when the media is played, specifically highlighting the exact information. Adam may optionally click on the player’s user interface to get additional information about the AI tool used for altering the content, the date of alternation, and the URL of the original content  **One or more of the following variations:**   1. Portions of the audio of the speaker are only altered. 2. Portions of the audio of the background are altered. 3. Additional faked closed shots are added to the original video. 4. The original file has only audio. An AI-generated video is created for the audio using archived videos of other events repurposed and altered to this event. 5. The video is altered such that a new audience is added to the video. While in the original video, there is no shot of the audience, in the new altered video, specific people who may not have been present, are added to the audience. |
| **Categorization** |
| Marking the alteration |
| **Preconditions** |
|  |
| **Potential Requirements** |
| * Marking of the AI-generated/altered content in each ISOBMFF track   + Information regarding the tools used, the time interval, the reference to the track with original audio content, the date, and the party responsible for creation and/or alteration.   + Easily findable and detectable information.   + Allows referencing one or more regulation schemes and/or regional laws.   + Support of multiple languages.   + Support of custom fields based on the applicable regulations. |
| **Feasibility** |
|  |

# Streaming- Enhanced audio/video

|  |
| --- |
| **Name:** |
| Improved audio of streaming multimedia |
| **Description:** |
| User Ada likes WWII history and watching news clips generated during that era from various websites. Since the files generated in that area often do not have good audio-visual quality, some service providers use AI technologies to enhance the audio tracks, especially when people are talking in outdoor environments.  Ada selects a video stream and watches the content. The stream contains an audio track that is enhanced using AI tools for better recognition of voices and clarity of words. The streaming player first warns Ada that the default audio track is enhanced using AI tools, and whether Ada wants to listen to that track or the original track. Furthermore, if Ada chooses the enhanced audio track, whenever the audio is enhanced, the player displays an overlay “enhanced quality of speech with AI”. Ada can click on the player user interface and listen to the original audio tracks, if she chooses, switching between two audio tracks during streaming.  **One or more of the following variations:**   1. Portions of one audio track are only enhanced. 2. Portions of the audio background track are suppressed. 3. Portions of the audio background track are boosted. 4. Additional closed shots are added to the original video. 5. The original file has only audio. An AI-generated video is created for the audio using archived videos of other events repurposed and altered to this event. |
| **Categorization** |
| Marking the alteration for enhancement |
| **Preconditions** |
|  |
| **Potential Requirements** |
| * Marking of the streaming manifest regarding AI-generated/altered content in each relevant track * Information regarding the tools used, the time interval, the reference to the track/adaptation set with original content, the date, and the party responsible for creation and/or alteration. * Easily findable and detectable information. * Allows referencing one or more regulation schemes and/or regional laws. * Support of multiple languages. * Support of custom fields based on the applicable regulations. |
| **Feasibility** |
|  |

# Streaming - Deep fake content

|  |
| --- |
| **Name:** |
| Deep fake content in multimedia streaming |
| **Description:** |
| User Adam receives emails about new news clicks. When Adam clicks on the link, a streaming manifest is downloaded and a multimedia player is launched to stream the content for Adam. The manifest describes a video track and audio track of 10 mins, showing a government official at a news conference. The video and audio tracks are altered from min 5 to min 7:30, fabricating ‘deep fake’ comments that were never made by the official. The player detects the AI-generated/altered marking in the manifest and displays a warning that the media content contains fabricated AI-generated content. If Adams decides to play the content, the player further signals the AI altered marks on the samples in both audio and video tracks, from min 5 to min 7:30 by overlaying the “AI altered content” over the content. Adam may optionally click on the player’s user interface to get additional information about the AI tool used for altering the content, the date of alternation, and the URL of the original content  **One or more of the following variations:**   1. Portions of the audio of the speaker are only altered. 2. Portions of the audio of the background are altered. 3. Additional faked closed shots are added to the original video. 4. The original file has only audio. An AI-generated video is created for the audio using archived videos of other events repurposed and altered to this event. 5. The video is altered such that a new audience is added to the video. While in the original video, there is no shot of the audience, in the new altered video, specific people who may not have been present, are added to the audience. |
| **Categorization** |
| Marking the alteration. |
| **Preconditions** |
|  |
| **Potential Requirements** |
| * Marking of the AI-generated/altered content in the manifest for each representation that is AI-altered. * Information regarding the tools used, the time interval, the reference to the track/adaptation set with original content, the date, and the party responsible for creation and/or alteration. * Easily findable and detectable information. * Allows referencing one or more regulation schemes and/or regional laws. * Support of multiple languages. * Support of custom fields based on the applicable regulations. |
| **Feasibility** |
|  |

# General requirements

The following requirements can be deduced from the above use cases:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Requirement | 3.1 | 3.2 | 3.3 | 3.4 |
| 1. Mark the ISOBMFF file as containing AI generated/altered content | X | X |  |  |
| 1. Mark the ISOBMFF track as AI altered. | X | X |  |  |
| 1. Mark one or more time intervals of the ISOBMFF track as AI altered. | X | X |  |  |
| 1. Mark the adaptation set containing AI-generated/altered content. |  |  | X | X |
| 1. Mark content components containing AI-generated/altered content |  |  | X | X |
| 1. Mark one or more time intervals of a period as AI altered. |  |  | X | X |
| 1. Easily finable information, i.e. the marking information to be discoverable with minimum parsing of the manifest or files, with no need for complex tools and standards. | X | X | X | X |
| 1. Support of multiple languages | X | X | X | X |
| 1. Nature of the alteration | X | X | X | X |
| 1. Date of the alteration | X | X | X | X |
| 1. Owner/responsible party for alteration | X | X | X | X |
| 1. One or more regulations that the marking is conforming to | X | X | X | X |
| 1. Additional information and fields per regulation | X | X | X | X |
| 1. The URL and/or track/adaptation set/representation of original content | X | X | X | X |