Forensic Attribution of TikTok Video Provenance Through Metadata and Re-Encoding Analysis

Domenico Valentino Lo Faro[1], Luca Guarnera[1,*] and Sebastiano Battiato[1]

[1] *University of Catania, Department of Mathematics and Computer Science, Italy*

Abstract

The rapid spread of manipulated and AI-generated multimedia content on social platforms increases the need for reliable forensic methods to verify the provenance and authenticity of digital evidence. TikTok, currently the most widely used platform for short-form video creation and sharing, performs several internal re-encoding, normalization, and metadata processing steps, which leave distinctive and potentially identifiable traces on exported video files. In this work, we conduct a systematic analysis of thirty case studies involving video acquisition, upload, and download from TikTok across heterogeneous devices, operating systems, and recording conditions. Using ExifTool and MediaInfo, we extract and compare metadata from original, uploaded, and downloaded files, while ensuring file integrity via SHA-256 hashing on multiple devices. Our results reveal stable forensic markers—including the Comment, Artwork, Information, Encoder, VendorID, VidMD5, and AigcInfo tags—as well as consistent re-encoding behavior across the platform, particularly the recurring use of FFmpeg libraries. These markers allow distinguishing native content from content that has passed through TikTok, enabling provenance attribution, detection of re-uploads or manipulations, and reconstruction of digital timelines. The findings provide actionable insights for investigators, forensic analysts, and practitioners dealing with social-media–based digital evidence.

Keywords

Digital Forensics, TikTok Provenance, Video Metadata Analysis, Re-encoding Analysis, Social Media Forensics, File Integrity Verification

1. Introduction

In the current socio-cultural context, the analysis and collection of digital evidence in judicial proceedings has become indispensable. In this field, a central role is played by social platforms such as TikTok, designed for the publication and sharing of short videos, generally in vertical format, which can be edited but also created directly through numerous filters, effects and an integrated music library (Figure 1).

This activity entails exposure to several risks, such as harmful content, the creation of addiction and the manipulation of the perception of reality, especially for younger users. Moreover, AI-generated content accounts for an increasingly large share of the media available on this platform, with all the risks that this implies, starting with the growing spread of fake news, compounded by the public's lack of preparation in distinguishing what is true from what is false. Several methods have been proposed to detect this manipulated and synthetic content [1, 2, 3, 4, 5, 6, 7, 8, 9, 10].

At the same time, TikTok's user base is constantly expanding and shows no sign of slowing down. It is estimated that by 2029 it will reach 2.35 billion users worldwide, with a growth rate of 21.6% recorded between 2018 and 2029 [11, 12].

The research presented in this article aims to develop technical and methodological criteria that can be applied in the forensic field for the recognition and traceability of content, contributing to the verification of the authenticity and origin of such content in investigative, journalistic or judicial

1st Workshop on Digital Forensics (WDF 2026), co-located with the 2nd Joint National Conference on Cybersecurity (ITASEC & SERICS 2026), February 09–13, 2026, Cagliari, Italy

*Corresponding author.

domenicovalentino.lofaro@studium.unict.it (D. V. L. Faro); luca.guarnera@unict.it (L. Guarnera); sebastiano.battiato@unict.it (S. Battiato)

https://www.dmi.unict.it/lguarnera/ (L. Guarnera); https://www.dmi.unict.it/~battiato/ (S. Battiato)

0000-0001-8315-351X (L. Guarnera); 0000-0001-6127-2470 (S. Battiato)

© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-14

CEUR-WS.org/Vol-4329/paper2.pdf

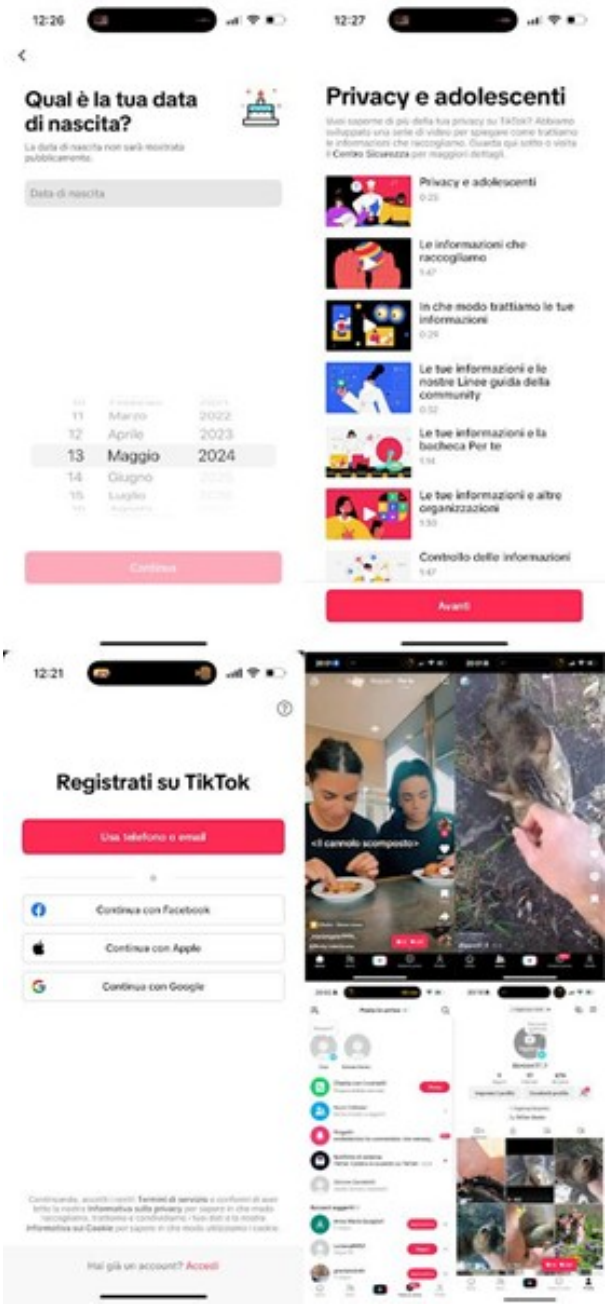


Figure 1: Example of a TikTok video interface used for recording and publishing short-form content. The platform provides numerous filters, effects, and editing tools that can influence the structure and metadata of the resulting video file.

contexts. The goal is to verify the presence, within video files, of features that demonstrate that they have passed through TikTok, in an attempt to identify trends shared by videos recorded on different devices, in different conditions and with different recording options. Throughout all stages of the research, systems and precautions were adopted to demonstrate the authenticity of the files analysed and the correctness of the information extracted. At the heart of the analysis are the metadata associated with the various files under examination. These were compared with the files obtained both after upload and after download. Through comparative analysis of video samples from different sources, specific patterns were identified in the metadata such as the tags comment, vid_md5, encoder, aigc_info, software which can constitute reliable indicators that a file has passed through the TikTok or Douyin ecosystem.

This work provides the following key contributions:

- Systematic experimental analysis involving thirty case studies across multiple devices, operating systems, and conditions (native recordings, TikTok recordings, edited videos, AI-generated content).
- Validation of file integrity across transfers using SHA-256 hashes on independent machines, ensuring the authenticity of all analyzed files.
- Identification of stable forensic markers—including Comment, Artwork, Information, Encoder, VendorID, VidMD5, and AigcInfo—that consistently indicate TikTok provenance.
- Characterization of TikTok's re-encoding pipeline, showing systematic use of FFmpeg libraries (e.g., Lavf58.76.100) in downloaded files and consistent normalization of file format and metadata.
- Definition of practical forensic indicators to distinguish original content from TikTok-transcoded content, supporting digital investigations, provenance reconstruction, and detection of re-uploads or manipulated files.

The remainder of this paper is organized as follows. Section 2 provides an overview of prior forensic studies on social network traces and the detection of manipulated/synthetic media. Section 3 describes the dataset, devices, tools, and methodology, followed by detailed results on the evolution of metadata during the recording, uploading, and downloading phases. Finally, section 4 provides an interpretation of the results, forensic implications, practical applications, and directions for future research.

2. Related Work

The impact that social media platforms have on the internal structure of uploaded images has been investigated since the mid-2010s. A seminal contribution in this area is the 2015 study by Moltisanti et al. [13], Image Manipulation on Facebook for Forensics Evidence. The authors show that the Facebook upload pipeline introduces systematic alterations to JPEG images, most notably a change in resolution. Their experiments reveal that the platform applies resizing rules based on the length of the longest side and on the selected upload mode (High Quality or Low Quality), following a deterministic scaling process that can be exploited for forensic purposes.

Another relevant finding concerns the renaming mechanism applied by Facebook after upload. The new filename consists of three parts: two random segments and a central Photo ID, from which valuable information can be extracted, such as internal references related to the image. The study also highlights that the Exif metadata of the original file are typically removed during the upload process.

The investigation of how social networks modify multimedia content was further expanded in 2017 by Giudice et al. [14]. This study analyses the behaviour of several platforms — including Facebook, Google+, Twitter, Tumblr, Flickr, Instagram, Imgur, Tinypic, WhatsApp and Telegram — and shows that most of them systematically alter image filenames. With the exception of Google+, these renaming patterns follow identifiable structures that may include useful elements such as URLs or internal identifiers for retrieving the uploaded file. These results contributed to shaping a methodological foundation for social network image ballistics, supporting the development of automated systems for attributing the origin and processing history of images circulating on social media.

Although our study focuses specifically on TikTok provenance, similar methodological principles appear in other domains. Rundo et al. [15, 16] show that mathematical modeling and pattern analysis can reveal consistent behaviors in complex systems such as financial markets, while Battiato et al. [17] demonstrate how stable perceptual features enable reliable visual content discrimination.These works highlight a shared paradigm—the extraction of robust, repeatable indicators—that also underpins our forensic approach to identifying systematic metadata and re-encoding traces left by TikTok.

3. Analysis

In the research presented in this article, multiple mobile devices were used for the recording, upload and download of the video files under study. This was done to ensure the absence of bias that could have resulted from the exclusive use of a single device.

As shown in Table 3, four mobile devices were used, two with the Android operating system and two with the iOS operating system.

The actual analysis of the videos was carried out on two desktop devices (Table 3), both running Windows.

The Asus VivoBook Pro 17 was used for the main video analysis, while the Lenovo ThinkPad was used to verify the authenticity of the video files after the transfer from mobile to desktop devices. The first software tool used for the analysis was PowerShell (version 7.5.1.0), used to extract the unique hash strings for each video file. The second tool was ExifTool (version 13.19 / 64-bit). The last tool used was MediaInfo (version 24.12), an open-source application. The videos captured on mobile devices were transferred to the desktop devices via standard USB cables. At this point, however, a problem arises: demonstrating that the transfer from mobile to desktop does not alter the integrity of the files under analysis, since any changes due to the transfer would compromise the research results, rendering them irreparably distorted. To obtain reasonable assurance as to the integrity of the videos, hash codes obtained via PowerShell were used: A hash code is a numerical value, computed through a hash function, used to represent an object's data in a compact form [1]. Hash codes are commonly used in data structures such as hash tables to optimise search, insertion and deletion operations. A hash code therefore is a string that uniquely represents an object, whether an image, a video or any other type of file. Any modification to the object for which the hash code is calculated will also change the hash code itself. PowerShell was used to extract hashes from the video files, which employs the SHA-256 algorithm.

It is clear that different objects have different, unique hash codes. In our analysis this property was leveraged to check the integrity of video files after transfer, by calculating these codes on two different machines(the VivoBook Pro 17 and the Lenovo ThinkPad) and verifying that on both the code remained the same. In both cases, therefore, the video file is identical and no changes occurred as a result of the transfer. The research is based on the analysis of metadata: structured information that describes, explains or facilitates the use of other data. Metadata can be divided into three main categories:

- Descriptive metadata: used to identify and retrieve digital objects, including information such as title, author, keywords and creation date;
- Administrative and management metadata: used to provide details on the management and preservation of documents, such as access rights, file format and retention policies;

[1]Microsoft "Ensuring Data Integrity with Hash Codes", 1 marzo 2023 https://learn.microsoft.com/en-us/dotnet/standard/security/ensuring-data-integrity-with-hash-codes

Table 1
Mobile devices used for recording and uploading the videos analyzed in this study. The table reports manufacturer, operating system, model number, and TikTok version, ensuring heterogeneity across platforms and acquisition conditions.

Device	Manufacturer	Operating system	Model number	TikTok version
iPhone 16 Pro Max	Apple	iOS 18.4.1	MYX03QL/A	39.7.0
iPad Air2	Apple	iPadOS 15.8.4	MGHY2TY/A	39.6.0
Galaxy A12	Samsung	Android 12	SM-A125F/DSN	39.8.4
Redmi 12C	Xiaomi	Android 13 TP1A.220624.014	22126RN91Y	39.8.4

Table 2
Desktop systems employed for metadata extraction, hashing, and video analysis. These devices were used to verify file integrity across transfers and to ensure reproducibility of the forensic evaluation workflow.

Model	Manufacturer	Operating system	Version
Vivo Book Pro17	Asus	Windows 11 Home	24H2
Thinkpad	Lenovo	Windows 10 Pro	2009

- Structural metadata: used to describe the internal structure of documents, indicating the relationship between their parts, such as chapters or sections.

- ExifTool and MediaInfo were used for metadata extraction, so that the results could be cross-checked using two different programs.

The analysis is based on the comparison between the "originals" and their respective "upload" and "download" versions. "Upload" refers to the video that TikTok saves on the device during the upload process. A total of thirty case studies were analysed, corresponding to twenty-six videos divided into six groups. The number of case studies does not exactly match the number of videos because in the fourth group (made up of four case studies) previously used files were reused.
Each "original" video is associated with one "upload" and one "download" video. The upload file is always located on the same device on which the original was recorded, while the download is always saved on a different device. Each group of videos has characteristics that differ from the others, in order to cover as many varied cases as possible:

1. **First group:** recorded and uploaded from Samsung Galaxy A12 and downloaded from Xiaomi Redmi 12C. The videos in this group were all recorded using the device's camera app and then uploaded to TikTok. No effects were applied.

2. **Second group:** recorded and uploaded from iPhone 16 Pro Max and downloaded from iPad Air 2. The videos in this group were all recorded using the device's camera app and then uploaded to TikTok. Audio effects were applied to videos 7, 9 and 10.

3. **Third group:** recorded and uploaded from iPad Air 2, Xiaomi Redmi 12C and iPhone 16 Pro Max, downloaded from iPhone 16 Pro Max and iPad Air 2. Videos 13 and 16 were recorded directly through the TikTok app. Audio effects were applied to videos 12, 13 and 16.

4. **Fourth group:** in this group, the objective is to compare videos downloaded from TikTok before and after sharing via WhatsApp. Videos 17 and 18 were transferred via WhatsApp Web, while the others were transferred via the mobile app.

5. **Fifth group:** the videos in this group were generated using artificial intelligence. Uploads were performed using iPhone 16 Pro Max, while downloads were done via iPad Air 2.

6. **Sixth group:** the videos in this group were all recorded and uploaded from iPhone 16 Pro Max and downloaded from iPad Air 2. Both audio and video effects from the TikTok app were applied to all of them.

In general, the upload files show a decrease in quality compared to the originals, which also affects the file size, which overall is reduced.

The first metadata tags to be analysed are Major Brand, Minor Version and MIME:

- Major Brand declares the file's "base" format. Through this parameter, the various playback or editing software tools are able to correctly interpret the file, indicating which extensions of the standard it supports and with which players/decoders it is most likely to be compatible[2].

[2]ISO/IEC 14496-12:2015 – Information technology – Coding of audio-visual objects – Part 12: ISO base media file format

- Minor Version is a 32-bit field (often expressed in hexadecimal form) that specifies the "version" or "revision" associated with the major brand[3].

- MIME indicates the file's content type according to the MIME (Multipurpose Internet Mail Extensions) standard. It is used to identify the format of the data contained in the file so that the operating system, browser or application knows how to handle it.

Original videos have different values for these tags depending on their nature, the device used to create them, and whether artificial intelligence software was used.
Upload files created on Android devices show the following values:

- Major Brand: MP4 Base Media v1 [ISO 14496-12:2003]
- Minor Version: 0.2.0
- MIME Type: video/mp4

In uploads from Apple devices, the tags take on two different sets of values:

- 1.
  - Major Brand: Apple QuickTime (.MOV/QT)
  - Minor Version: 0.0.0
  - MIME Type: video/quicktime
- 2.
  - Major Brand: qt
  - Minor Version: 512
  - MIME Type: video/quicktime

The difference in tag values for uploads from iOS devices is only superficial, since the two sets of values are identical. In the first case, the metadata extraction software (ExifTool) provides a textual description, while in the second case qt is the official identifier value of the QuickTime File Format, registered at mp4ra.org. In addition to being saved on the same device from which they are uploaded, upload files also retain the same extension as the original videos, i.e. .mp4 for Android devices and .MOV for iOS. The values of the Major Brand, Minor Version and MIME tags reflect this difference.
These results suggest that the re-encoding performed during upload assigns the same standard values to the MIME, Major Brand and Minor Version tags, keeping differences only when the device's operating system, and therefore the video upload format, changes. Another noteworthy tag is Artwork. It is present in all videos uploaded from iOS devices, but not in those uploaded from Android (although other unique tags, discussed below, are present there). Usually the Artwork tag indicates the presence

[3]ExifTool documentation : https://exiftool.org/exiftool_pod.html

of images embedded in the file. In all videos analysed, however, it contains descriptive information on the content and its origin in the form of a JSON string. When no effects are applied to the video, the value of this tag is:
-Artwork:


```
{"data":{"product":"tiktok","editingEffectsList":[]}}
```


We can see that it unequivocally links the video in which it appears to the passage through TikTok: ("data":{"product":"tiktok"}). In the videos in the sixth group, i.e. those where visual effects and filters were applied during upload, the JSON string is different:
-Artwork :


```
{
  "data": {
    "product": "tiktok",
    "editingEffectsList": [
      {
        "selectedFrom": "editing_effect_template",
        "subType": "edit_prop",
        "timeRangeList": [
          {
            "endTime": 27.379168,
            "startTime": 3
          }
        ],
        "effectName": "Masked Arab",
        "effectId": "2066512927",
        "containAIGCEffect": false
      },
      {
        "selectedFrom": "editing_effect_template",
        "subType": "edit_prop",
        "timeRangeList": [
          {
            "endTime": 3,
            "startTime": 0
          }
        ],
        "effectName": "Eye blue beautiful",
        "effectId": "1498340977",
        "containAIGCEffect": false
      }
    ],
    "prop_list": "MjAzMTI5Mjg2OSwxNDE2ODU5NDg4"
  }
}
```


The Information tag, as already mentioned, appears in almost all videos uploaded from Android and has a JSON text as its value:
-Information :


```
{
  "data": {
    "prop_list": [],
```

```
    "met_publish_data": [
      {
        "effect_list": "",
        "effect_source": "",
        "is_parent_template_used": false,
        "met_parent_resource_ids": "",
        "prop_list": "",
        "prop_selected_from": "",
        "resource_list": "",
        "tab_key": ""
      }
    ],
    "product": "aweme"
  }
}
```


This value is shared by all videos, again uploaded from Android, to which no effects were applied. When the videos are edited, the Information tag contains details about the effects used:
-Information:


```
{
  "data": {
    "prop_list": [],
    "met_publish_data": [
      {
        "effect_list":
           "269095,269096,269097,18635115,82238160,85085370,85085370,18635115,
            90786170,82238160,99915446",
        "effect_source": "",
        "is_parent_template_used": false,
        "met_parent_resource_ids": "",
        "prop_list": "",
        "prop_selected_from": "",
        "resource_list": "",
        "tab_key": ""
      }
    ],
    "product": "aweme"
  }
}
```


In the Information tag, as in Artwork tag for videos uploaded from iOS, we find a direct link to TikTok through the field "product":"aweme". “Aweme” is the technical (backend) name of Douyin, i.e. the version of TikTok available in China [18]. In many cases where the upload was performed from an Android device, the Information tag is accompanied by a Description tag. However, the presence of the latter is not as consistent: it is missing from the third, twenty-eighth and twenty-ninth videos. Like Information, this tag also has a JSON text as its value:
-Description


```
[
  {
    "make": "",
    "videoIndex": 1,
    "userSystem": "12",
```

```
    "appRecord": "0",
    "MD5": "39fb8276b2cf146782c10a22ee45382b",
    "videoDevice": "",
    "system": "12",
    "userDevice": "SM-A125F",
    "deviceResolution": [],
    "importPath": "content://media/external/video/media/1000030986",
    "videoResolution": {
      "width": 864,
      "height": 1920
    },
    "videoDuration": 15660,
    "isRecord": 0,
    "isCropped": 1
  }
]
```


In all videos except the eighth, eleventh, twelfth, thirteenth, sixteenth and all those in the fifth group, the Copyright tag appears:
-Copyright:a712067981afe469174e298a5c5f25c6
It does not indicate a legal licence such as "© TikTok 2024" but rather a 32-character hexadecimal string which probably identifies the file within a database. As already mentioned, almost all videos uploaded from Android have the Information and Description tags. Whenever these tags are present, the Encoder tag takes the value "bytevehwavc". This value is not associated with any official standard, so it is hypothesised that it is a string representing a proprietary encoder belonging to ByteDance, the Chinese company that owns TikTok. The only exception to what has been said so far is the third video of the first group which, despite being recorded and uploaded from an Android device, it shows the Encoder tag with the value "Lavf57.71.100": "Lavf" is the abbreviation of "libavformat" [9], one of the FFmpeg libraries, while "57.71.100" indicates the version of this library. With regard to videos uploaded from iOS, in cases where the upload was done from iPhone 16 Pro Max, two tags identifying the software used for recompression are found: Encoder with the value "Lavf57.71.100" and Vendor ID with the value "FFmpeg". In the videos of the third group uploaded from iPad Air 2 (namely the eleventh, twelfth and thirteenth) only the Vendor ID tag with the value "FFmpeg" appears. In the fifth group, whose videos were generated using artificial intelligence, video 21 contains both the Vendor ID tag with the value "FFmpeg" and the Compressor Name tag with the value "Lavc59.37.100 libx264". The prefix Lavf, as seen, indicates the use of the libavformat library belonging to FFmpeg. The remaining videos in the fifth group show only the Vendor ID tag with the value "FFmpeg".

All videos recorded in vertical orientation have a Rotation tag with a value of 90°, whereas in the upload files it is 0°. This indicates that the mobile device physically records the video in landscape and then applies a 90° rotation to make it appear vertical, while TikTok saves the file directly in vertical orientation without applying any rotation. Uploading to TikTok leads to reassignment of temporal metadata. At video level, all upload files show a frame rate of about 30 fps, while the audio-related tags indicate the presence of an AAC LC stereo audio track. AAC LC stands for Advanced Audio Codec Low Complexity. It is one of the most common audio standards used with both .mp4 and .mov files (ISO/IEC 13818-7:2006). Furthermore, all upload files have a sampling frequency of 44.1 kHz and a bitrate of 128 kbps. As already observed for uploads, downloads also show a reduction in file size (even more pronounced than that seen for uploads). This is consistent with the lower quality of downloads compared to the originals. Unlike uploads, the format of downloads does not depend on the device from which the upload (or download) is performed. Instead, it is always .mp4 on both iOS and Android. Specifically, all downloads exhibit the same values for the Major Brand, Minor Version and MIME Type tags, which provide information on the file format:

- Major Brand: MP4 Base Media v1[ISO 14496-12:2003]
- Minor Version:0.2.0
- MIME Type: video/mp4

From the value “MP4 Base Media v1 [ISO 14496-12:2003]” we can deduce that the file is an MP4 in a base version, specifically v1, from which more recent versions such as v2 derive. It also shows that the video was encoded using the “ISO 14496-12:2003” standard. Taken together, these factors show that the downloaded file has been re-encoded into a format which, while not recent, provides high compatibility with the vast majority of devices regardless of their operating system and its version.All downloads, regardless of group, have a Comment tag. This tag appears to be specific to TikTok, as it is present in all case studies in this research. Comment is followed by a string. Here are some examples:

- Comment: vid:v24044gl0000cuqvklnog65ifrriou0g (Video1);
- Comment: vid:v24044gl0000cuqvlsfog65o6h2sqicg (Video2);
- Comment: vid:v24044gl0000cv2nnpfog65ot39prskg (Video6);

We can see that no two Comment values are identical. This implies that this tag is used as a unique identifier for the video both within TikTok’s servers and externally. It is also observed that all strings begin with the value “vid”: and that the initial part of each string is always the same in all cases: “vid:v24044gl0000”. It can therefore be hypothesised that this string works as a counter that increments every time a video is uploaded to the platform, which explains why only the final part changes. It is important to underline that the value of Comment does not appear to depend on the account from which the upload or download is performed, since different accounts were used in the various groups. Moreover, on iOS devices where the download took place, the file is saved using the Comment string as the file name minus the vid: prefix. For example, Video 6, whose Comment tag has the value:“vid:v24044gl0000cv2nnpfog65ot39prskg”, is saved with the name: “v24044gl0000cv2nnpfog65ot39prskg” .
Unlike uploads, in downloads the Encoder tag is constant and always takes the value “Lavf58.76.100”. As already mentioned, “Lavf” is an abbreviation indicating libavformat, an open-source library belonging to FFmpeg. We can therefore conclude that all downloaded videos have been re-encoded using FFmpeg, with not only the same library but also the same version, “58.76.100”. The Rotation tag, whose value is always 0° in downloads, leads to the same conclusions drawn for uploads. In some cases the Aigc Info and Vid MD 5 tags are present. They appear in downloads 8, 9, 10, 11, 21, 26, 27, 28, 29 and 30. Aigc Info may indicate the presence of AI-generated elements within the video. In all cases where this tag appears, it has the value “{"aigc_label_type":0}”. “{"aigc_label_type":0}” indicates the absence of AI-generated elements within the videos, whereas “{"aigc_label_type":1}” would indicate their presence. However, whenever this tag appears in this research it always has the value “{"aigc_label_type":0}”, indicating that TikTok does not detect the presence of AI-generated elements, despite the fact that all videos in the fifth group were generated using AI. The Vid MD 5 tag appears in the same downloads where Aigc Info is also found. The value of this tag is a 32-character hexadecimal string corresponding to an MD5 hash (128 bit). It may be interpreted as an additional “digital fingerprint” capable of uniquely identifying the video. All metadata relating to the location and to the device that recorded the original video are absent. The audio is stereo (two channels) with a depth of 16 bit and uses the AAC LC format, the same as observed for uploads, with a sampling frequency of 44.1 kHz. In addition, all downloads use CABAC: CABAC, acronym for Context-Adaptive Binary Arithmetic Coding, is an entropy coding algorithm used in video compression. In particular, CABAC uses binary arithmetic coding to represent data in an extremely compact way and adapts the coding process to the context of the data (that is, to the statistical characteristics of the bits being encoded), improving compression efficiency. It is employed in advanced video standards such as H.264/AVC and H.265/HEVC, where its greater compression efficiency justifies the increased computational complexity compared to other methods such as CAVLC (Context-Adaptive Variable Length Coding). In short, CABAC makes it possible to obtain smaller video files while maintaining high quality, at the cost of a greater computational load during encoding and decoding.

All downloads use multiple reference frames (four), whereas the original videos use only one. In video, reference frames are frames used as a basis for predicting subsequent frames during encoding. In other words, many video codecs (such as H.264, HEVC, etc.) do not encode each frame independently, but instead exploit information already present in previous frames (the reference frames) to represent only the differences relative to them. This method, known as inter-frame coding, makes it possible to reduce the amount of data to be transmitted or stored, thereby improving compression efficiency.

4. Discussion and conclusions

Cybercrimes-offences committed through the use of IT tools, received full international recognition with the Budapest Convention of 23 November 2001, later ratified by the Italian Parliament with Law No. 48 of 18 March 2008. More recently, Law No. 90 of 28 June 2024 further strengthened national cybersecurity and increased attention to the prevention of computer crimes, a phenomenon that is now widespread at both national and global level and in constant evolution. In this context, it is clear that TikTok, currently the largest social network, is increasingly becoming an epicentre of cybercrime and computer-related offences. The unlawful conduct that may arise is manifold and ranges from the distribution of illegal content to sextortion, child pornography and the grooming of minors, i.e. online solicitation through various social networks. Finally, content uploaded to TikTok can be highly relevant as evidence for identifying individuals involved in criminal acts and for reconstructing events. The repeated study of multiple video files generated under different conditions, as carried out in this research, makes it possible to recognise patterns and fixed tag values that allow a probable origin to be attributed to the files analysed in a forensic context, also determining whether a file has been transcoded, manipulated or re-uploaded. If a video exhibits metadata tag values typical of TikTok, such as those identified in this research, it is possible to conclude that:

- The file was not recorded directly on an original device, but has passed through TikTok;
- It may be a reuse or manipulation of content that has already been published.
- This information is useful in:
- Judicial proceedings in which it is necessary to prove the provenance of the content;
- Investigations into deepfakes, disinformation or revenge porn, where the origin of the file is crucial.

In forensic contexts where it is necessary to analyse large volumes of data, as may occur for example following the seizure of a smartphone, the presence of known patterns can speed up the process of determining the origin of files. Moreover, knowing the origin of the files under examination can help reconstruct a timeline of digital activities, show that a piece of content was disseminated and not merely created, and link social-media activity with the files found on the device.

Acknowledgments

The work has been supported by the Spoke 1 "Future HPC BigData" of the Italian Research Center on High-Performance Computing, Big Data and Quantum Computing (ICSC) funded by MUR Missione 4 Componente 2 Investimento 1.4: Potenziamento strutture di ricerca e creazione di "campioni nazionali di R&S (M4C2-19)" - Next Generation EU (NGEU).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] L. Guarnera, O. Giudice, S. Battiato, Deepfake Detection by Analyzing Convolutional Traces, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 666–667.

[2] S. Concas, G. Perelli, G. L. Marcialis, G. Puglisi, Tensor-Based Deepfake Detection in Scaled and Compressed Images, in: 2022 IEEE International Conference on Image Processing (ICIP), IEEE, 2022, pp. 3121–3125.

[3] J. Battocchio, S. Dell'Anna, A. Montibeller, G. Boato, Advance Fake Video Detection via Vision Transformers, in: Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, 2025, pp. 1–11.

[4] A. Ciamarra, R. Caldelli, F. Becattini, L. Seidenari, A. Del Bimbo, Deepfake Detection by Exploiting Surface Anomalies: The Surfake Approach, in: 2024 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), 2024, pp. 1024–1033. doi:10.1109/WACVW60836.2024.00112.

[5] L. Guarnera, O. Giudice, S. Battiato, Mastering Deepfake Detection: A Cutting-Edge Approach to Distinguish GAN and Diffusion-Model Images, ACM Transactions on Multimedia Computing, Communications and Applications (2024).

[6] U. A. Ciftci, I. Demir, L. Yin, FakeCatcher: Detection of Synthetic Portrait Videos Using Biological Signals, IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2020).

[7] E. Coletta, D. Salvi, V. Negroni, D. U. Leonzio, P. Bestagini, Anomaly Detection and Localization for Speech Deepfakes via Feature Pyramid Matching, in: Eurasip EUSIPCO, 2025.

[8] S. Concas, S. M. La Cava, R. Casula, G. Orrù, G. Puglisi, G. L. Marcialis, Quality-based artifact modeling for facial deepfake detection in videos, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 3845–3854.

[9] S. Concas, S. M. La Cava, A. Panzino, E. Masala, G. Orrù, G. L. Marcialis, Deceptive Beauty: Evaluating the Impact of Beauty Filters on Deepfake and Morphing Attack Detection, arXiv preprint arXiv:2509.14120 (2025).

[10] L. Guarnera, O. Giudice, S. Battiato, Level Up the Deepfake Detection: a Method to Effectively Discriminate Images Generated by GAN Architectures and Diffusion Models, in: Intelligent Systems Conference, Springer, 2024, pp. 615–625.

[11] https://www.castmagic.io/post/tiktok-user demographics, Tiktok users: Age insights and helpful statistics, 2024.

[12] N. Dunn, Top 23 tiktok statistics & facts you need to know in 2025!, 2025.

[13] M. Moltisanti, A. Paratore, S. Battiato, L. Saravo, Image Manipulation on Facebook for Forensics Evidence, in: International Conference on Image Analysis and Processing, Springer, 2015, pp. 506–517.

[14] O. Giudice, A. Paratore, M. Moltisanti, S. Battiato, A classification engine for image ballistics of social data, in: International Conference on Image Analysis and Processing, Springer, 2017, pp. 625–636.

[15] F. Rundo, F. Trenta, A. L. di Stallo, S. Battiato, Grid Trading System Robot (GTSbot): A Novel Mathematical Algorithm for Trading FX Market, Applied Sciences 9 (2019) 1796.

[16] F. Rundo, F. Trenta, A. L. Di Stallo, S. Battiato, Advanced Markov-Based Machine Learning Framework for Making Adaptive Trading System, Computation 7 (2019) 4.

[17] S. Battiato, G. Gallo, S. Nicotra, Perceptive Visual Texture Classification and Retrieval, in: 12th

International Conference on Image Analysis and Processing, 2003.Proceedings., volume 3, 2003, pp. 524–529. doi:10.1109/ICIAP.2003.1234103.
[18] P. Lin, TikTok vs Douyin A Security and Privacy Analysis, Technical Report, 2021.