Synthetic media is exploding: can metadata help to keep the digital space transparent?

The last year has shown us that synthesised videos are becoming a new media reality. The alternative Christmas speech by Queen Elizabeth and deepfake Tom Cruise TikTok videos generated by artificial intelligence and machine learning technologies are vivid examples that such technology has the potential for viral proliferation.

The spike of AI and ML integration into the digital world is leading to numerous concerns about the credibility of content. Users cannot blindly trust content creators and publishers. It becomes important to understand the source of the material and whether it has been digitally manipulated in some way. The scope of this question about digital origins is increasing exponentially. The number of deepfakes doubled over the previous year and reached *85,047* videos in December 2020.

Given this trend, it's becoming increasingly important to to build transparency in the digital medium. This involves a means to mark content with information on its creator, origin, and recent changes – this can be done with an old friend to digital files – like metadata.

Metadata and its value

Metadata is a reference to data about other data – with synthetic media, it's shorthand information about a digital asset. Today, metadata gives essentially a snapshot about a particular file. To understand how it works, think about visiting the library. When you want to know about a book, you look at it's descriptive information – the author, title, date of publication, publisher, etc.

But there's another tool to learn more about that book's journey, which is recorded in a library card. It also contains the borrowing history. The library card adds information beyond its intrinsic characteristics. You get an additional layer of information that provides additional insights on its journey before it reaches your hands.

And there's work underway to transform metadata so it will work the same way. You will be able to add attribution to a picture or video so that other content consumers can check its origin and track previous changes.

This gives transparency that can help any consumer of media content understand all of the same steps in a media file's journey before they encounter it. This is especially important for synthetic media, where it's not always clear if a file is digitally manipulated – at all – and how. Without knowing that background, the content could be used in manipulative and even threatening ways.

Stakeholders are creating a regulatory basis

Metadata development to add these layers of information about a piece of content's history for synthetic media is a process that is just beginning. A *Deepfake Task Force Act* proposed earlier this year offers a coordinated plan that explores how a digital content provenance standard can reduce the threat of deepfakes. However, it's not the only document that coordinates synthetically originated content in the digital space.

In 2019, *Adobe* introduced *the Content Authenticity Initiative* (CAI) in collaboration with the *BBC* and *Washington Post*.

CAI members recognize that it is crucial to expand cooperation between partners and introduce a universal standard for other companies and social media. The digital world needs regulations for platforms to identify content uploaded and shared with metadata. That, in turn, will contribute to companies

actively enrolling and coordinating content in the global digital ecosystem.

Getty, Microsoft, The New York Times, Twitter, as well as synthetic media companies like Synthesia and Reface, have already *joined* CAI. These companies are working to map out solutions regarding content attribution and changes performed to the content.

Challenges on the way to data attribution

In order to develop the theoretical ground to putting metadata into practice, there are several major challenges regarding a creator's privacy, copyright, and data storage.

1. Creator data loss

Lack of information about the source of origin is one of the biggest challenges today. Still, many independent creators are actually unable to associate their name to generated content in a way that doesn't get lost or don't do it at all. Hiding behind anonymity is a straight road to an unsecured digital space, specifically when trust in digital content is constantly falling. And it happens all the time in the phase between when content travels from the author to consumers.

Once unattributed images and videos are published on the web, no one can track any possible modifications to that content. It, thus, requires significant effort to detect whether it's a deepfake or not.

2. Threat to content authenticity

Metadata, in turn, can support copyright protection and boost user awareness of the consumed digital content. For example, Tom Cruise's *fans were shocked* to find out a series of impressive videos on TikTok starring another person "deepfaked" as the real Tom Cruise was not actually the Hollywood actor. This prompts the question for every new video of Tom Cruise released since February: *was it created with AI and ML technologies* or is it really Tom?

The synthetic version is so good, it's virtually impossible to tell the difference between the actor

and his copy on the screen.

And, at present, there is not a built-in way to check whether you're looking at real actors and events, or digitally manipulated ones on the various distribution platforms themselves. But even if you had the raw files – it might also not be immediately clear if what you're consuming is authentic or not. So there is an opportunity to mark files so that everyone can be aware of its origin.

This case may also serve as a warning for businesses, opinion leaders, politicians, and celebrities, because content created with their likenesses could have important ramifications. This is why it's crucial to anticipate data management *within digital business processes* that can also eliminate copyright controversies.

3. Adopting a single standard

Last but not least, it's crucial to build an integrated and unified system with formal review standards and capabilities for storing data about digital content. For example, once deciding to make or modify a video or image, every user should understand that this content requires identification before being shared on the web.

It's possible to store pertinent information about a creator while only displaying select information and guaranteeing their privacy. Users would have to register and attribute their username to their generated content, just like on TikTok and some other social media platforms.

While enhancing metadata isn't a cure-all, it may be an effective tool to solve some of the struggle with deepfakes. While several companies are working on authentication and validation tools, the introduction of updated metadata standards can also be a realistic and promising solution. This technology is capable of meeting the needs of all stakeholders – creatives, business representatives, and content consumers – while also finding a level of control to address malicious uses of AI/ML-generated videos.

Anna Bulakh is Policy Advisor at *Reface*

Article by ANNA BULAKH