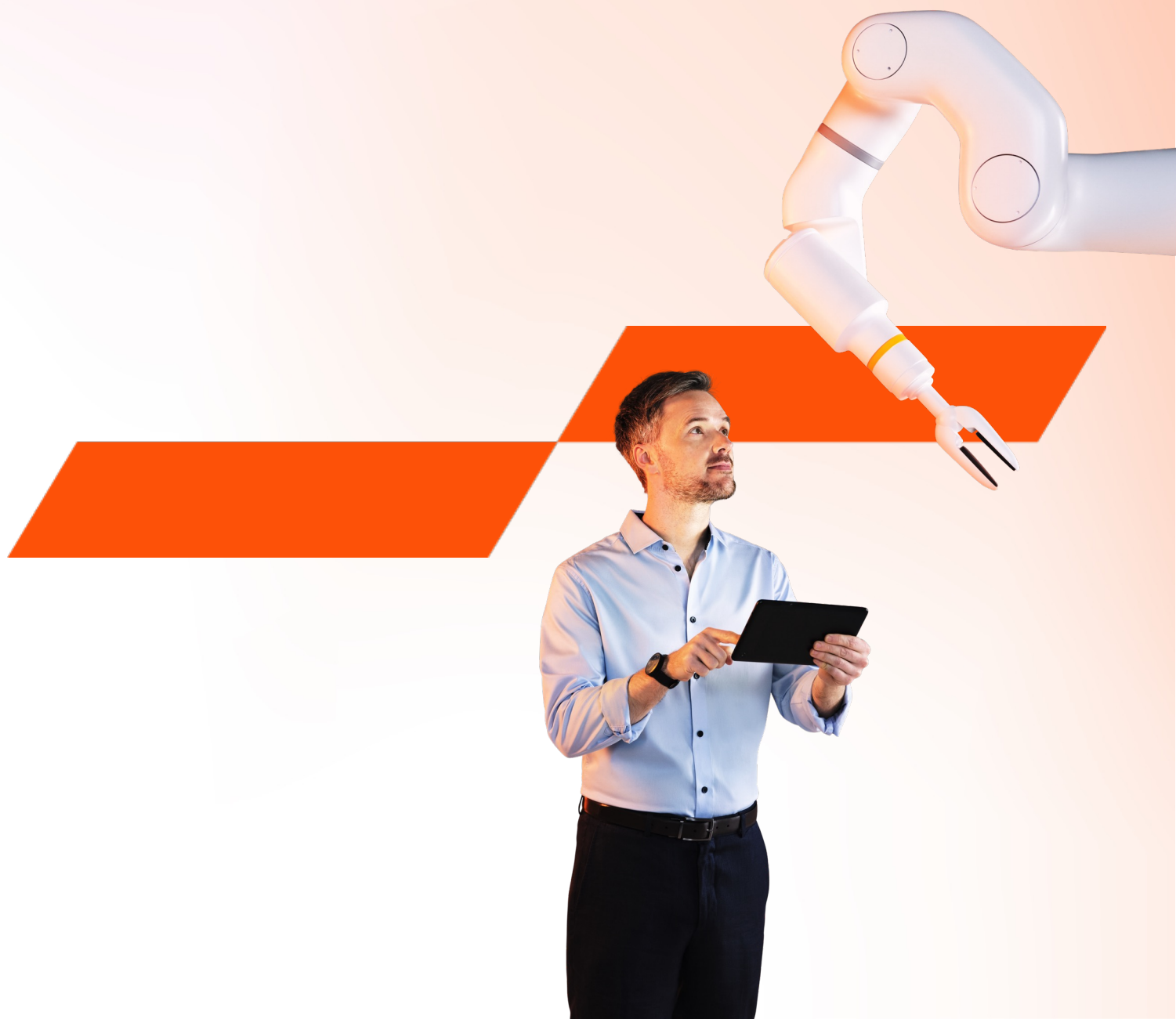




The future of digital trust: Why content provenance matters more than deepfake detection





Introduction

The evolving nature of digital trust

A few years ago, people assumed that a video, photograph, or audio recording can be taken as evidence of reality. Whether someone said something in a video conference, gave a statement or an instruction, he or she was considered real and the content was validated as being true as well.

Unfortunately, this is not the case anymore.

Due to the rapid development of generative AI technologies, it has become possible to generate and manipulate images, voice recordings, and videos—making them look authentic. These synthetic media assets, known as deepfakes, are now widely accessible through publicly available and easy-to-use platforms.

Challenges for the communications and media industry

The communications and media industry thrives on trust and portraying authentic content. All news, entertainment, messages from the government, brand statements, digital campaigns, etc., are based on an underlying trust that there is a credible source behind each piece of media.

Due to the rampant deepfakes penetrating the internet, the consequences of this new threat go far beyond the technical layer. In addition to trust, organisations' reputation, business decisions, public trust, etc. are also being affected.

For organisations, the question, 'Can we detect fake content?' becomes secondary to the question, 'Can we prove that the content is genuine?'

The emergence of deepfakes

Deepfakes represent media content (images, videos, voice) generated with AI, imitating real-life objects and people. This type of media assets can include videos with edited faces, cloned voice recordings, fake photos, fake interviews, or real-time impersonation on video calls.

In the communications and media industry, the threat of deepfakes is higher due to the rapid pace of the industry. For instance, a fake video can easily spread across digital platforms in minutes. Even if it is detected and refuted later, the damage would already be done.

Deepfakes can:



Spread misinformation during critical events



Impersonate business leaders



Damage brand reputation



Create false endorsements



Manipulate public opinion



Commit fraud



Bypass identity checks



Create confusion during crisis situations

What is concerning is that this technology is developing very fast. Earlier, deepfakes often had visual problems—unnatural lip movements, bad lighting, unusual facial expressions, or robotic voice. Nowadays, most of these issues are hardly noticeable.

A person watching a short video clip on a smartphone may not be able to recognise whether the content is genuine or manipulated.

This means that an organisation cannot rely only on the human ability to detect manipulated content.



Why detection is not enough

Many companies start tackling the deepfake issue with detection-based solutions. Such tools analyse media files to detect any traces of manipulation. This includes things like facial movement, audio, lighting, artifacts of image compression, or unusual metadata.

Detection is useful and would continue to be a vital component in securing digital media. However, the problem cannot be solved through detection alone since there is more that needs to be understood about deepfake technologies that have grown increasingly advanced due to generative AI's development.

The first limitation in detection-based systems is their reactive nature. Most of the time, detection happens once the media file has been made and released to the public. In an increasingly rapid technological world, such content can be spread very quickly through social media, news outlets, and messaging services before it is discovered. Consequently, the damage done to one's reputation, finances, or society as a whole might take place before it can be controlled.

Secondly, detection technology engages in a never-ending race with generation technology. While AI models become more advanced in their ability to create realistic media files, detection models must stay competitive and evolve accordingly. The winners often remain temporary in this scenario.

Third, is confidence. Sometimes a detection tool would conclude that the video is 'probably fake' or 'potentially synthetic', but this is not enough for certain decisions to be made. Companies need something more for certain situations.

Detection asks: Is this fake?

Authentication asks: Can we prove where this came from and whether it has been altered?

Authentication: A stronger basis for digital media protection

Authentication aims to build trust starting from the point of content creation. Instead of waiting for suspicious content, authentication records all the necessary information about this content throughout its lifecycle.

For example, when a picture is taken, authenticator systems could record where it was taken, which device was used, by whom, if it was edited or modified, and how. Such a record could accompany the content all the way from creation, to editing and finally to publishing.

Simply put, authentication creates a chain of trust in digital content.

This is a more sustainable approach compared to detection since it does not rely exclusively on identifying things that are fake. Rather, it helps to establish the authenticity of the content.

From media companies' perspective, this approach could be very useful. A newsroom, broadcaster, digital publisher, or communications department could prove to their audience that their content is authentic and unaltered.



The importance of content provenance standards

An important development in addressing digital trust and deepfakes is the emergence of content provenance standards.

Content provenance is the history of creation and distribution of a digital asset including its origin, modification, and delivery to the end viewer.

One of the initiatives in this space is the Coalition for Content Provenance and Authenticity (C2PA). C2PA defines the standards for attaching content credentials to digital media.

These content credentials can be compared to a product label which includes information about a product's origin, production process, and expiration date.

Similarly, content credentials include information about a digital file, including aspects such as:



Author of the content



If it was edited



Date of creation



Tools used in the process



Whether AI technologies were involved



Status of content credentials (was it modified?)

It's important to note that the introduction of the content provenance standards doesn't mean that all content will be considered credible. However, these standards play a vital role in giving stakeholders a chance to make decisions about the credibility of the content based on the information provided in the content credentials.



Deepfake threats for the communication sector

The communications and media industry is at the core of this challenge since it produces, curates, distributes, and amplifies digital content.

- A deepfake video with an influential person can shape the public debate.
- A forged message can ruin the trust of consumers.
- A cloned voice of a company's CEO can mislead employees and partners.
- A deepfake news can cause confusion during a crisis.

For media companies, trust is not only a compliance matter but a business asset.

Lack of trust in the authenticity of digital content is going to have a negative effect on the entire content value chain. Journalists might need extra verification processes. Platforms might need special labelling tools. Advertisers would require evidence of authenticity of the content they are going to use in their campaigns.

The ability to verify the authenticity of digital content will soon become a competitive advantage for media organisations.

Recommendations for organisations

01

Build the trust within the content lifecycle

Organisations should focus on building trust instead of detecting manipulations with digital content.

02

Focus on high-risk use cases

All images and videos don't require verification processes right now. Organisations can start with high-risk use cases like executive communication, financial disclosures, statements to the public, news assets, and crisis communication.

03

Adopt provenance standards

Media and communications organisations should look at the adoption of content credentials and provenance-based verification standards.

04

Educate employees and viewers

Employees need to understand how to deal with deepfake attacks and how to verify unusual requests. Also, the audience should understand the concept of content credentials and authenticity labels.

05

Preserve human accountability

AI tools can help with detecting manipulations and verifying content, but the ultimate responsibility should stay with human analysts.

Conclusion

Generative AI is transforming how trust is built online. In the case of the communications and media sector, it involves more than just trying to detect manipulated material—it also requires building confidence in the validity of information being disseminated to audiences and stakeholders.

As synthetic media continues to develop rapidly, organisations need to move away from a purely reactive strategy based on detecting manipulated content and adopt a proactive strategy centred on trust, transparency, and accountability with authentications, provenance, and content validation techniques into the content process.

While one technology alone will never be able to solve the problem of deepfakes, there are new standards being developed—like content credentials and provenance—which can form the bedrock of a more trustworthy and robust online ecosystem. Companies that will embrace these capabilities today will be better prepared for the future.





About PwC

We help you build trust so you can boldly reinvent

At PwC, we help clients build trust and reinvent so they can turn complexity into competitive advantage. We're a tech-forward, people-empowered network with more than 370,000 people in 149 countries. Across assurance, tax and legal, deals and consulting we help build, accelerate and sustain momentum. Find out more at www.pwc.com.

PwC refers to the PwC network and/or one or more of its member firms, each of which is a separate legal entity. Please see www.pwc.com/structure for further details.

© 2026 PwC. All rights reserved.

Contact us

Debankur Ghosh

Partner, Emerging Businesses
PwC India
debankur.ghosh.in@pwc.com

Harish Kalyanaraman

Director, Emerging Businesses
PwC India
harish.kalyanaraman@pwc.com

Shruti Solanki

Senior Manager, Emerging Businesses
PwC India
shruti.x.solanki@pwc.com

Contributor

Udayakumar Dama

Editor

Rubina Malhotra

Design

Kanagavathi Murugaiah

Data Classification: DC0 (Public)

In this document, PwC refers to PricewaterhouseCoopers Private Limited (a limited liability company in India having Corporate Identity Number or CIN : U74140WB1983PTC036093), which is a member firm of PricewaterhouseCoopers International Limited (PwCIL), each member firm of which is a separate legal entity.

This document does not constitute professional advice. The information in this document has been obtained or derived from sources believed by PricewaterhouseCoopers Private Limited (PwCPL) to be reliable but PwCPL does not represent that this information is accurate or complete. Any opinions or estimates contained in this document represent the judgment of PwCPL at this time and are subject to change without notice. Readers of this publication are advised to seek their own professional advice before taking any course of action or decision, for which they are entirely responsible, based on the contents of this publication. PwCPL neither accepts nor assumes any responsibility or liability to any reader of this publication in respect of the information contained within it or for any decisions readers may take or decide not to or fail to take.

© 2026 PricewaterhouseCoopers Private Limited. All rights reserved.

KA/August 2026 – M&C 54654