MULTI-STAKEHOLDER MEDIA PROVENANCE MANAGEMENT TO COUNTER SYNTHETIC MEDIA RISKS IN NEWS PUBLISHING

J. Aythora[1], R. Burke-Agüero[2], A. Chamayou[2], S. Clebsch[2], M. Costa[2], J. Deutscher[2], N. Earnshaw[1], L. Ellis[1], P. England[2], C. Fournet[2], M. Gaylor[2], C. Halford[1], E. Horvitz[2], A. Jenks[2], K. Kane[2], M. Lavallee[3], S. Lowenstein[3], B. MacCormack[4], H. Malvar[2], S. O'Brien[1], J. Parnall[1], Elissa M. Redmiles[2], A. Shamis[2], I. Sharma[2], J.W. Stokes[2], S. Wenker[2], A. Zaman[2]

[1]The British Broadcasting Corporation, United Kingdom, [2]Microsoft Corporation, USA, [3]The New York Times Company, USA, [4]CBC/Radio-Canada, Canada

ABSTRACT

The rise of indirect content distribution via third party social media platforms has introduced a new conduit for synthetic or manipulated content. That content purports to be legitimate news, or to come from legitimate news sources, and can present the consumer with apparent brand integrity markings, which convey authority.

Three major global news organizations and a leading technology provider have come together to demonstrate a mechanism to tackle this problem that can operate at scale. The BBC, The New York Times Company, and CBC/Radio-Canada in cooperation with Microsoft have developed a proposed open standards approach which can be used by large and small news organizations to protect the provenance of news stories in audio/visual/textual media.

INTRODUCTION

The rise of social media and video hosting platforms has created a significant problem for identifying content provenance on the internet. Re-hosting of media has meant that the origin of media content is increasingly obfuscated, undermining consumer trust and enabling the propagation of dis/misinformation[1] often using established and trusted brand imagery to amplify the deception.

[1] We use the term disinformation to cover the broadest definition of information disorder – disinformation, misinformation and malinformation

In order to meet this societal challenge, it is important to consider both technical and media business perspectives. Consequently, the authors of this paper have come together to demonstrate a provenance verification system that can be implemented at massive scale.

Our approach enables consumers to determine the publication source of media, independent of the site or server hosting it. This will foster trust in the *provenance* of the media, and offer assurance that media is authentic and has not been altered since its original publication.

We will present a prototype implementation of an open-standards media provenance architecture. This has been developed to enable content publishers to authenticate content as part of their publication workflow and for consumers to verify the content as received. The paper will detail the components of this architecture, including media provenance registration, provenance data binding to the media, provenance data distribution and consumer verification.

The system architecture has been developed to support many types of publishers and media, including streaming video. We envisage this initial implementation will provide the stimulus for wider standardization of the common interoperable data structures and interfaces required, leading to a distributed ecosystem of content provenance system implementations and operators.

SCOPING THE DISINFORMATION THREAT

Disinformation – A multi-faceted problem

Disinformation can enter the news ecosystem in many forms. First Draft has defined seven types of mis and disinformation [1]. This paper will address the risks caused by **Imposter Content**, **Manipulated Content, Mis-contextualization** and **Fabricated Content**.

Our aim is to authenticate the provenance and status of a piece of media by technically linking it to its published source and signaling any tampering in its distribution*. We do not make any assessment of the relative truth or trustworthiness implicit in it, or that of the publishing organization or reporter.*

Deep Fakes and Brand Hijacking – The next generation of threat

The problem of malicious actors assuming the trusted brand identities of well-known news publishers is a current reality. Media now often reaches its audience via indirect paths, independent of the publishers'/broadcasters' own digital sites. The malicious use of the established brand markings allows bad actors to add credibility to fictitious works. With the advent of AI generated Deep Fakes, there is now a risk that powerful traditional symbols of authority, trusted news brand hosts and sets, can be used to amplify disinformation.

There are three approaches that can be deployed to counter the risk of Deep Fake synthetic content in news. The first is **Media Education**. This relies on training the consumer to increase their level of skepticism. While effective, it runs counter to decades of effort to build audience trust in news brands. The second approach is to use AI-based Deep Fake **detection algorithms**. This may have only short-term efficacy. Generative Adversarial Network implementations can use these tools to recursively test and improve the sophistication of the fakes they are meant to detect. This leaves **Provenance** as the third defensive strategy. Provenance strategies, to be effective, will require a coordinated approach across the news publishing, social media and technology eco-systems. This is the reason the Origin Alliance was formed.

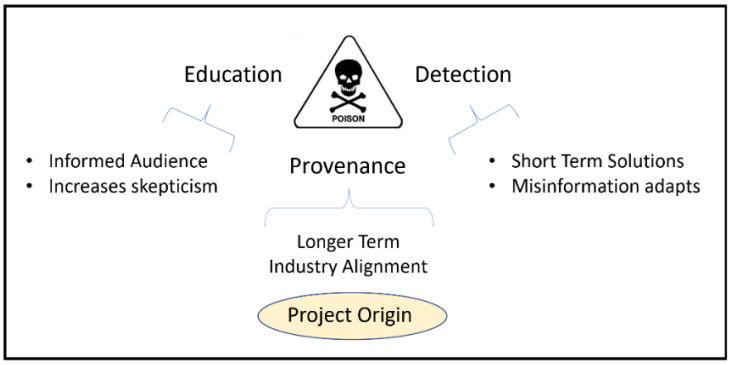

Figure 1 - Three Responses to Deep Fake News

PROVENANCE – THE AUTHENTICATION OF MEDIA

Separation of the signal from the noise

As the amount of disinformation continues to add noise to news ecosystems, it becomes important to have a consistent method to easily identify valid signals. Adding provenance information and binding it to the media amplifies information coming from publishers and broadcasters, making valid signals and therefore trustworthy content, easier to identify.

Special attention will need to be taken during the design to allow for early deployment cases. Initially, very few legitimate news sources will have implemented the Origin system, and the player will need to reflect a neutral, rather than negative opinion on provenance.

There will also be cases where the provenance of media needs to be intentionally obfuscated for the security of the reporter. The system needs to allow for a recognized actor to attest to the source of provenance without providing detailed information.

The chain of provenance

News items are built using multiple inputs. The intention of the Origin approach is to build a chain of provenance from the point of publishing to the point of presentation. The news publisher, via their news standards and practices, will attest to the provenance of all upstream sources. Other work is being done (by Adobe and others [2]) to capture the provenance chain from the lens to the editorial system. Conversations to create an end-to-end open standard are ongoing.

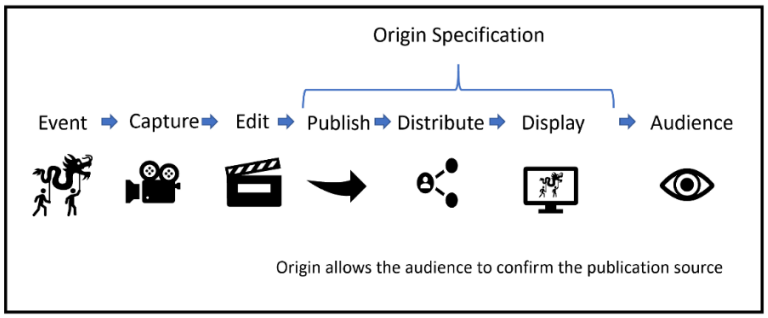

Figure 2 - The Media Provenance Sequence

Diverse publishing environments and formats

News publishers vary in scale, technical capacity and media types employed in storytelling. An effective provenance solution must be accessible and affordable for all producers of news content. The larger global media brands involved in the creation of the Origin Alliance recognize that any system implemented will need to have a way of being simply implemented by news organizations of all sizes. Access to a positive (authenticated) provenance system cannot become a barrier to entry for any news organization.

It is also important to emphasize that positively determining the provenance link between a media story and its publisher is in no way an editorial endorsement of the validity of the news content. The Origin consumer user interface is intended to show provenance; care must be taken to distinguish provenance from trustworthiness (or truth), which is a wider and more complex issue.

The techniques of authentication will vary across the different digital formats for text, audio, photo and video files. The embedding techniques will vary by format, however, the data structures for provenance should be common. The Origin approach defines a minimum common data requirement, with extensions for media type and for variations between publishers.

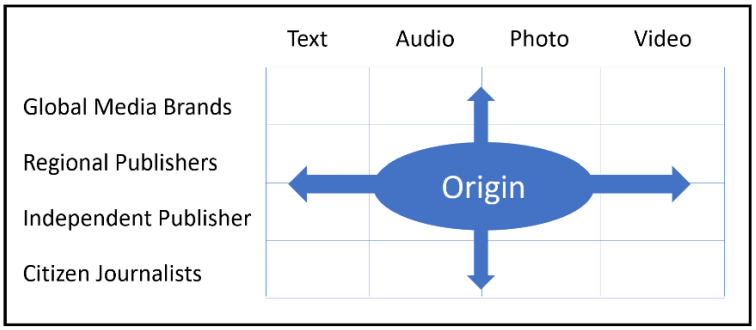

Figure 3 - An Extensible Provenance Standard

Participation across the ecosystem

In addition to alignment between news publishers, a fully effective provenance solution will require cooperation across the complete technology stack. Cloud media services and editorial tool vendors will have to offer common feature implementations. Social media platforms will have to monitor for provenance signals and provide appropriate distribution treatments based on validity of the response. This will be a larger industry conversation. Many of these discussions are underway via the Partnership on AI [3] and its Media Integrity group [4].

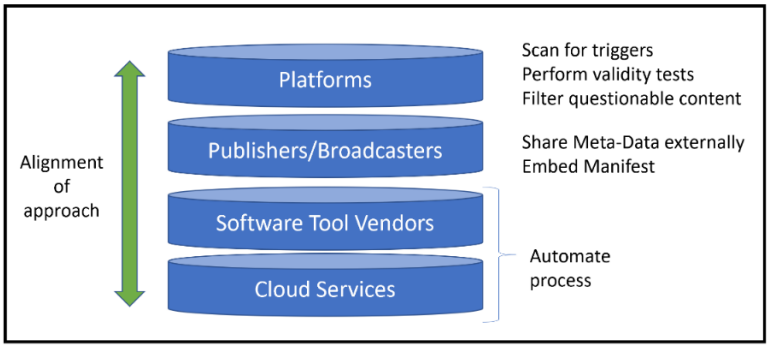

Figure 4 - Cross Industry Alignment

ORIGIN – A PROPOSED MEDIA PROVENANCE SOLUTION

Organization

Members

- **Founders** - The four groups that originally came together to articulate the problem, the proposed approach and advocated for wider industry acceptance. BBC / NYT / CBC/Radio-Canada & Microsoft
- **Contributors** - Organizations who join with the founders to contribute human, financial and/or technical resources and know-how to advance the design and adoption of a common solution.
- **Adopters** - Organizations that will work to support and adopt the industry solution as it goes through the standardization process and is adopted by a wider range of users.

Alliance Approach to the disinformation problem

- We will not try to prove the nature of a manipulation or to spot general deep fakes in the wild. Detection and analysis of synthetic media is a separate problem.
- We are not concerned with authentic publisher content shared legitimately or harmlessly either by partners or third parties. Origin is not a digital rights management solution.
- We want to create something that works throughout the media ecosystem. We are open to cooperation with other publishers and technology partners.
- We will develop solutions with the goal of creating open standards. Provenance solutions require broad participation in a common approach to be effective.
- We believe whatever we build will be one part of a wider solution to the problem of disinformation, but not solve the problem in its own right.

TECHNICAL CHALLENGES ON DISTRIBUTION

When media is distributed at scale over the public internet, it typically undergoes several modifications outside of the publisher's control as it travels from publication to consumption. These modifications lead to changes to the binary content of the source file. Robustness to such benign modifications, while still detecting malicious attacks, is the core goal of our work.

Traditional linear broadcast network distribution and publisher controlled digital sites are not within the scope of our work, as these systems are well controlled from source to destination.

Benign Modifications – Positive Authentication

For efficient bandwidth utilization in Internet links, media is commonly distributed over content distribution networks (CDNs). A key function of modern CDNs is re-hosting; CDNs can store source media in several locations across the world, so final media streaming/delivery is done close to the clients. Re-hosting often comes with a change of origin, particularly in scenarios where the CDN function is run by a media distributor (such as Akamai).

In many cases CDNs will decode and re-encode (i.e. transcode) media, to optimally adjust the media quality (resolution and bit rate) to the assessed network characteristics, connecting a particular device and the chosen media resolution. Such transcoding can be minor, such as re-quantization within the same codec (coder/decoder) format, or it can be substitution, such as re-encoding on a different format. Such re-encoding may be on the fly or based on cached versions of different qualities and resolutions, as it occurs with adaptive streaming.

Transcoding presents a challenge to provenance tracking approaches such as hashing, because the actual bits in the new media file are different from the original media. Another challenge is that certificates and other ancillary data carried on metadata fields of the media format may be also lost during transcoding.

Our technical solution considers two core distributions models. First, where CDNs are enhanced to comply with the proposed provenance tracking approaches, so that cryptographically-secure certificates and hashes are preserved in the metadata of transcoded files. Second, where CDNs and other benign content modifiers are not under control, and metadata could be lost. In those cases, we will leverage fingerprinting and watermarking techniques, so the media content itself carries provenance information, even after transcoding.

Malicious Attacks – Negative Authentication

Attacks take the form of edits to the source media that are intended to deceive or mislead.

Deep fakes [5] are an example of an attack intended to alter consumers' understanding of the provenance or the nature of the media they are viewing. A deep fake might use a previously shot and published video, but manipulate the words being spoken. It may also modify the faces of persons in the video, to synthesize a video of a person of interest appearing to be making a statement that they never did.

Shallow fakes are another kind of attack, where small modifications to media are made that change how that media is perceived, in a way that misleads or manipulates. An example is manipulating frames to slow down a video or editing/permuting/inserting frames in ways that change the context of the media. Shallow fakes can technically be very similar in their edits to the benign edits above, and as such present a challenge in understanding which edits are "authorized" and which are not.

Any provenance solution should not provide authentication to such maliciously altered media.

Indeterminate states during system rollout – Neutral Authentication Signal

There will be times when no judgement on the provenance media can be determined. During the initial deployment period this will be the majority of the cases. During this early period, it is important that the lack of full provenance information is not intended to convey distrust in the content. A neutral signal will represent an unknown provenance determination, consistent with current distribution practice.

TECHNICAL ASPECTS OF THE ORIGIN PROJECT

The Origin Project has been our joint effort to develop a robust and scalable solution for media provenance certification. Its roots are in three initially separate projects by the four alliance partners, with common approaches to the problem of media verification, including the New York Times News Provenance Project (NPP) [6], the BBC/CBC Provenance Project, and the Microsoft AMP system. The external groups became aware of each other via discussions, many of which sprang from Partnership on AI sponsored events – notably a convening in London in May 2019. We are now combining the efforts to build a common approach, with a shared intention of developing an open industry standard.

Although the three previous independent efforts were seeking to address a separate solution to media provenance, each was focusing on distinct solutions. Origin combines those ideas in a system that we believe will lead to short- and long-term success.

As shown in Figure 5, there are three main aspects of the Origin system:

1) a Media Provenance Service
2) the consumer user experience research
3) usage scenarios and ecosystem

We briefly address each one next.

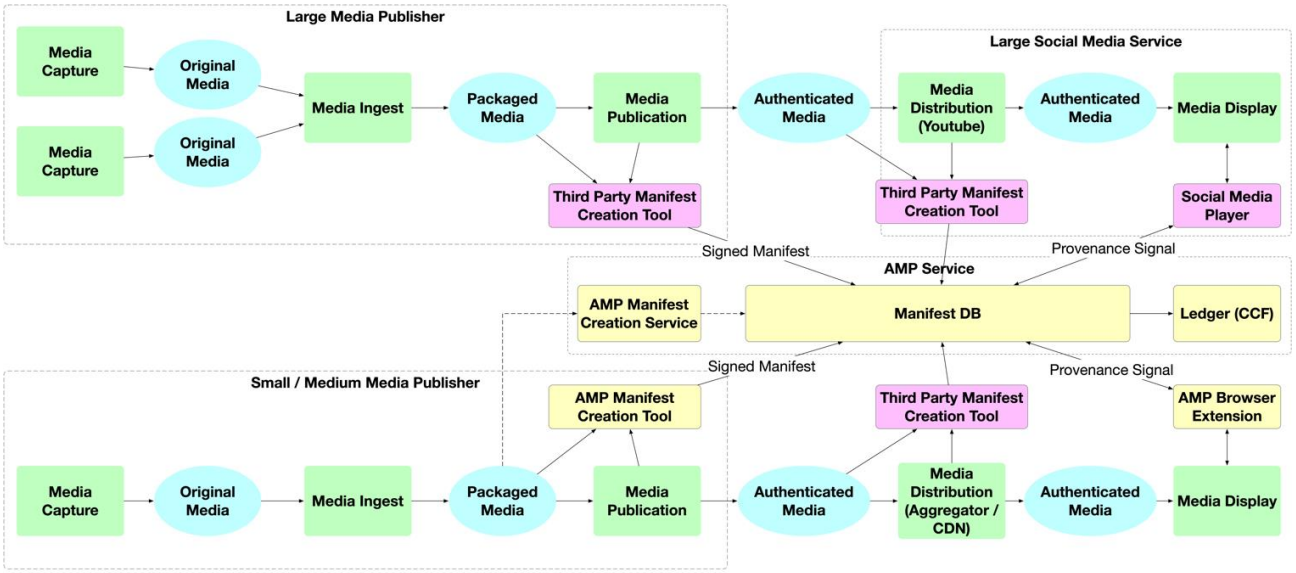

Figure 5 - Origin Project Ecosystem Overview

AMP Service

The Origin Project uses the AMP System [7] to implement the provenance service. Microsoft Research and the Azure Media Services team proposed and implemented the AMP system, which seeks to *A*uthenticate *M*edia via *P*rovenance. The key components of the AMP Service are highlighted in yellow in Figure 5.

AMP includes two separate media provenance service modules: 1) A module that provides a near-term solution, with little or no modifications to the underlying media transport and display subsystems, and 2) a module that provides a long-term media provenance solution, but requires more extensive changes to both the media and W3C browser standards, as well as the actual Web browsers themselves.

The AMP design focuses initially on video but is also applicable to images and audio. A primary goal of AMP's near-term solution is to provide media provenance with no or minimal changes to the media transmission and consumption pipeline. At its core, a publisher creates a signed *manifest*, including the hash for the media data. For video, for example, a streaming manifest uses hashes of chunks of the media data, for providing authentication for adaptive streaming. The publisher then uploads the media's manifest to the AMP service which uses a database to store these manifests. Clients such as social media applications or Web browsers can then query by the provenance information and manifests from the AMP service. Manifests and their cryptographic hashes are registered on a ledger, which is tamper proof and provides strong resiliency guarantees across multiple hosts. One approach is to use the ledger technology in the *Confidential Consortium Framework* (CCF) [8]. A description of the ledger system and the manifest specification and implementation can be found in [7]. We foresee Origin using CCF for its initial distributed ledger.

For robustness in distribution scenarios that include benign modifications to the media content, AMP will have provisions to use watermarking and fingerprinting technologies to embed the manifest within the media content itself [7]. If malicious modifications are made to the content during distribution, the AMP service can detect the absence of valid watermarks from modified media.

AMP's near-term provenance module focuses on little or no changes to media infrastructure. It also has a second module, which seeks a long-term solution exploring more robust solutions that involve possible modifications or extensions to media creation, transport, and rendering standards.

In our first prototypes to evaluate the AMP system, we have included extensions to the streaming manifest, so that the provenance signal can be generated within the HTML video element itself. The AMP prototype also includes a modified version of the Edge Chromium browser, to display the provenance signal within the browser itself.

User Experience

A very important aspect of the Origin Project is media consumer user experience. Origin's approach to user experience design has been driven by ideas from the three separate initiatives:

- A provenance-based browser extension implementation and user research conducted by the BBC and CBC/Radio-Canada partnership.
- A browser extension for the AMP service developed by Microsoft
- An extensive set of user experience research studies authored by the New York Times News Provenance Project for browser extensions and web page authoring

Browser Extensions for Indicating Provenance

The BBC and Microsoft independently implemented browser extensions that can convey a provenance signal to the end user via a browser extension. In the BBC's provenance system, the provenance signal is created by comparing streaming audio fingerprints of the original media with the media being played in the Internet browser. AMP also includes a Chrome browser extension for displaying provenance information for videos played on YouTube. The AMP browser extension generates the provenance signal if the media's cryptographic hash matches the manifest stored in the AMP service. Both of these efforts contribute to the 'Media Display' section, below.

User Research

Origin's future user experience will be based on findings from the New York Times' News Provenance Project (NPP). A critical aspect of a media provenance solution is how to alert the end user that the media object that they are viewing can be authenticated back to its purported source. To address this issue, NPP explored how to build trust and confidence in online media by attaching context to a photograph as it travels around the Internet. Through user research and technical prototyping, NPP pointed to the importance of a set of common symbols and standards that give readers the tools they need to trust what they are seeing online. Metadata that is recognizable and familiar — like date, location, source, and related photos — offer more to build trust than unfamiliar process or technological descriptions. Establishing a common technical and UX foundation that works across media types will help to strengthen this sense of familiarity and trust in contextual information for readers and media producers. Project Origin will use the NPP research findings to drive the design of future browser extensions as well as providing prototype web sites which can used by designers to leverage the Origin media provenance signal in the design and implementation of their web sites.

ORIGIN USAGE SCENARIOS

Manifest Creation and Signing

A publisher first needs to create a manifest which is uploaded to the AMP service, and the method used to create the manifest of the media depends on the size of the publisher. Several use cases for Origin for both large and small/medium-sized media publishers are depicted in Figure 5.

Large-scale media producers operate their own internal production pipelines, which include state-of-the-art information technology (IT) divisions. Their media production pipelines are complex and automated. From the Figure, multiple Original Media objects may be captured by different reporters, photographers or videographers which are combined to create a Packaged Media object. One important requirement of these

large-scale media publishers is that the content of the Packaged Media may not necessarily be uploaded or stored in a cloud service. Therefore, they typically use a third-party signing tool as part of their Media Publication pipeline.

On the other hand, small- or even medium-sized publishers may not have an extensive automated media publication pipeline. Much of the work that is done to create the Packaged Media may be done manually. In some cases, the reporters may also help create the Packaged Media instead of automated software. To support this case, AMP provides two methods for small- and medium-sized publishers to sign their media and create a manifest which are both depicted in Figure 5. As with large-scale media publishers, a small- or medium- sized publisher may not want to upload their Packaged Media to a cloud-based service. In this case, the AMP system provides the AMP Signing Tool and the AMP Manifest Creation Tool which can be locally hosted. Alternatively, the publisher can use the AMP's Cloud Service to create the manifest, but this requires uploading the media to the cloud.

Media Distribution

During distribution, multiple entities including content distribution networks and media aggregators may modify the original authenticated media, and a manifest for these derived media objects must also be uploaded to the AMP Service. Similar to the media publishers, the manifests can be created locally by these distribution entities using the AMP Manifest Creation Tool.

Media Display

Finally, the media must be displayed to the end user typically by a social media application such as YouTube or Facebook, or by an individual web page using a browser. In the case of an application running on a mobile device or a computer, these applications can be modified to communicate with the AMP Service and consume the provenance signal which is displayed in the application's user interface. In this case, the application designer can benefit from the research provided by the News Provenance Project to better inform users about the provenance information associated with the media object which is being displayed.

For the case of the end users consuming content via a web browser, provenance information can also be provided by the AMP Browser Extension, although the programming paradigm provided by browser extension APIs lead to much less intuitive results for the user.

CONCLUSION

The Origin Alliance partners have identified the risk that synthetic and manipulated news information presents to the ability of publishers to maintain audience confidence. Embedding a secure provenance tracing process into the collection and distribution system provides an enduring solution. This approach will require broad cooperation and horizontal alignment across the publishing industry and vertically across the technology and distribution stack. This must be done with sensitivity to the differences in economic and technical resources of the those who wish to use it. The increase in confidence provided with secure provenance cannot be achieved at the risk of silencing voices.

The approach presented in this paper uses new applications of well-established technical trust mechanisms. The Origin provenance system is meant to securely convey and protect the pre-existing editorial trustworthiness of organizations, journalists, and content. That trustworthiness is an input to the system.

The deployment of an effective provenance system will be as much organizational as technological. It is the intention of the authors act as a focal point for the organizational conversations that will be required. The creation of an open industry standard, and its widespread adoption, is the goal of the alliance. We look forward to working with others to bring this effort to fruition.

REFERENCES

1. C. Wardle, “Fake news. It’s complicated”, First Draft, Feb. 2017. Available at: https://firstdraftnews.org/latest/fake-news-complicated.

2. “Setting the industry standard for digital content attribution”, The Content Authenticity Initiative, 2019. Available at: https://contentauthenticity.org.

3. The Partnership on AI. Available at: https://www.partnershiponai.org.

4. “AI and Media Integrity Steering Committee”, The Partnership on AI, 2019. Available at: https://www.partnershiponai.org/ai-and-media-integrity-steering-committee.

5. G. Barber, “Deepfakes Are Getting Better”, WIRED Magazine, May 2019. Available at https://www.wired.com/story/deepfakes-getting-better-theyre-easy-spot.

6. “News Provenance Project”, New York Times, 2018. Available at: https://www.newsprovenanceproject.com/resources.

7. P. England *et. al.*, “AMP: Authentication of Media via Provenance”, arXiv:2001.07886, Jan. 2020. Available at https://arxiv.org/abs/2001.07886.

8. M Russinovich e.t al., “CCF: A Framework for Building Confidential Verifiable Replicated Services”, Microsoft Research Technical Report MSR-TR-2019-16, Apr. 2019. Available at https://www.microsoft.com/en-us/research/publication/ccf-a-framework-for-building-confidential-verifiable-replicated-services.