An Introduction to the JPEG Fake Media Initiative

Frederik Temmermans[1,2], Deepayan Bhowmik[3], Fernando Pereira[4] and Touradj Ebrahimi[5]
[1] Department of Electronics & Informatics (ETRO), Vrije Universiteit Brussel, Brussels, Belgium
[2] imec, Leuven, Belgium
[3] Division of Computing Science & Mathematics, University of Stirling, Stirling, FK9 4LA, UK
[4] Instituto Superior Técnico, Instituto de Telecomunicações, Lisboa, Portugal
[5] Multimedia Signal Processing Group, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
frederik.temmermans@vub.be, d.bhowmik@ieee.org, fp@lx.it.pt, touradj.ebrahimi@epfl.ch

***Abstract*—Recent advances in media creation and modification allow to produce near realistic media assets that are almost indistinguishable from original assets to the human eye. These developments open opportunities for creative production of new media in the entertainment and art industry. However, the intentional or unintentional spread of manipulated media, *i.e.*, modified media with the intention to induce misinterpretation, also imposes risks such as social unrest, spread of rumours for political gain or encouraging hate crimes. The clear and transparent annotation of media modifications is considered to be a crucial element in many usage scenarios bringing trust to the users. This has already triggered various organizations to develop mechanisms that can detect and annotate modified media assets when they are shared. However, these annotations should be attached to the media in a secure way to prevent them of being compromised. In addition, to achieve a wide adoption of such an annotation ecosystem, interoperability is essential and this clearly calls for a standard. This paper presents an initiative by the JPEG Committee called *JPEG Fake Media*. The scope of JPEG Fake Media is the creation of a standard that can facilitate the secure and reliable annotation of media asset creation and modifications. The standard shall support usage scenarios that are in good faith as well as those with malicious intent. This paper gives an overview of the current state of this initiative and introduces already identified use cases and requirements.**

***Index Terms*—JPEG, fake media, standardisation, media creation and modification, deepfake, authenticity, media forensics**

I. Introduction

Nowadays, media assets can be modified or even entirely synthetically created, *e.g.*, using deep learning methods, in such a way that they are hard to distinguish them from original media assets to the human eye [1]. These developments open new creative opportunities that are useful for the entertainment industry and other business usage, *e.g.*, creation of special effects, artificial but photo realistic scene production with actors in the studio or restoration/re-colourisation [2]–[4]. However, this also leads to issues relating to fake media generation defying the integrity (*e.g.*, deepfakes), copyright infringements and defamation, to mention a few. Misuse of manipulated media can cause social unrest, spread rumours for political gain or encourage hate crimes [5]–[8].

In many application domains, the creators may want, and even need, to declare the type of modifications that were performed on the media asset, in opposition to other situations where the intention is to hide the mere existence of made manipulations. In fact, it is important to dismount the myth that media modifications are always negative, *i.e.*, manipulations, as they are increasingly a normal and legal component of the production pipeline. This is already leading various governmental organizations to plan new legislation [9]. It is also true that companies, especially social media platforms and news outlets, are developing mechanisms that would clearly detect and annotate manipulated media when they are shared, attempting to avoid the negative impacts. While growing efforts are noticeable in developing technologies, there is a need to have a standardized way to annotate media assets (whatever the intent) and securely link them together. Therefore, the JPEG standardization committee (under hospices of ISO, IEC and ITU) has launched an initiative to identify the standardization needs related to the facilitation of the secure and reliable annotation of modified media through an in-depth analysis of various usage scenarios. While the initiative is called *JPEG Fake Media*, it is important to stress that it addresses both good faith and malicious usage scenarios. Therefore, in this paper, the term *Fake Media* is used to refer to any generated or modified media asset, independently of its '*good*' or '*bad*' intention, as well as media used in a misleading way.

It is expected that as a follow up of this effort, JPEG initiates a standardization activity in order to ensure interoperability between a wide range of applications dealing with fake media. To reach this goal, JPEG has and continues to invite stakeholders to join the effort by helping to better understand applications and scenarios relevant to fake media use cases. This allows the JPEG committee to identify key requirements for a standard in fake media. Initial findings suggest that a set of standard mechanisms to annotate fake media along with relevant information on the latter are needed. In addition, standard mechanisms for security and protection of integrity of media assets are desired. The latter is closely related to issues highlighted in media blockchain which has been under progress for a few years in JPEG [10] and therefore is considered as a natural continuation of that effort.

It is also important to understand, in more depth, the usage scenarios which will require input from relevant industries, public bodies (responsible for legislation), technology providers and end-users. Therefore the JPEG Committee engages with stakeholders in order to develop a clearly defined roadmap for standardization. In this context, this paper has

the objective of introducing the JPEG Fake Media initiative to the relevant academic, research and industry partners, in order to gather contributors to develop the best possible JPEG Fake Media standard.

II. Background

As stated previously, media modification is subject to usage scenarios both in *bad* and *good* intents. This section discusses some of the technological advances emerged in the recent past including in media forensics, and media creation and modification in creative industries.

A. Media forensics

Media forensics has become even more significant due to recent rise in spreading of fake news. Emergence of deep learning based manipulation, aka *deepfakes*, poses a new level of challenges in forgery detection, particularly with ready availability of software, *e.g.*, Face2Face [11], NeuralTextures [12] and FaceSwap[1], which can produce near realistic media content. The deepfake generation may involve use of generative adversarial networks (GAN). The scale of the problem prompted large tech companies like Google to create deepfake dataset [13] with over 1.8 million manipulated images for open research and multiple governmental bodies around the world to discuss policies for law enforcement. A handful approaches are available that use common deep learning training [13] on images or identifying the patterns for facial expressions and movements [14].

B. Media creation and modification in the creative industry

While *deepfakes* are commonly associated with of *bad* intention, *deep learning* techniques are, in fact, increasingly used in the creative industry from content creation to media restoration. Colourisation of archived media [4], [15] or black and white photographs [16] shows tremendous potential for the digitisation of pre-colour era photography or cinematography. Yoo *et al.* [3] successfully applied Convolutional Neural Network (CNN) models for coloring animated media / comic sketches using limited data. Among other usage scenarios, Virtual Reality (VR) is one of the emerging fields where deep learning is being used for content creation and exploration [2]. Similarly CNN have been of major use in media restoration [17]. Further in cinematography processing researchers and industries show the use of deep learning in video generation through scene dynamics [18], speech driven facial shape animation [19], image object animation [20], style transfer to character expressions [21], creating film trailer [22] or character motion synthesis and editing [23]. Recently, the TV show "For All Mankind" extensively uses deepfakes to bring characters including Johnny Carson, John Lennon, and Ronald Reagan back to live in an alternate history story line [24].

With such a diversity in usage scenarios, it is increasingly important that a standardisation effort could help content creators and other stakeholders including those who are interested in detecting media manipulation. To this end, the Content Authenticity Initiative[2], a consortium of industry bodies aims to "develop the industry standard for content attribution. By augmenting subjective judgments about authenticity with objective facts about how a piece of content came to be, the CAI aims to help content consumers make more informed decisions about what to trust" [25]. While the CAI focuses on the content authenticity framework, the focus of JPEG is on the signalling syntax for media. Both initiatives are closely working together to enforce each other in reaching their goals.

III. Definitions

To ensure a correct understanding of the terminology used in this paper, and to avoid any misunderstanding of concepts and issues raised, this section provides definitions of terms used in the context of this initiative. This will in particular to better clarify the objectives of the JPEG Fake Media initiative.

Fake Media: any generated or modified media asset, independent of its 'good' or 'bad' intention, as well as media used in a misleading way.

Media asset: digital assets in form of images, videos, audio or text including their combination; in the context of this initiative, the main focus is on images; however, other types of media are not necessarily excluded from the scope.

Media asset content: the content representation of the a media asset excluding metadata, for example the pixel data in case of images.

Media asset metadata: data associated with the media asset content, such as annotations or IPR information.

Media asset origin: the method or the device that was used to create the media asset.

Misinformation: false or inaccurate information that is communicated regardless of an intention to deceive.

Disinformation: a species of misinformation that is deliberately deceptive.

Modification: changes made to a media asset.

Manipulation: modification with the intention to induce misinterpretation.

Generated media asset: artificially created media asset.

Media asset integrity: lack of corruption of a media asset.

Authentic media asset: media asset that is verifiable and/or trustworthy

[1] github.com/MarekKowalski/FaceSwap/

[2] https://contentauthenticity.org/

Verifiable: able to be checked.
Trustworthy: able to be relied on as truthful.

Original media asset: media asset that has not been modified since its origin.

Provenance: the set of information about a media asset including (but not limited to) the media origin and trail of modifications.

IV. Use Cases

One of the key objectives of the JPEG Fake Media initiative is to better understand topics and use cases that fall under its scope and to analyse their implications, especially from a standardization point of view. At the time of this writing, the JPEG committee has identified the following topics and use cases:

- **Misinformation and disinformation**
  - Deepfakes
  - Manipulated media
  - Media intentionally used out of context
- **Forgery / Media forensics**
  - Document forgery (*e.g.*, IDs and passports)
  - Insurance fraud (*e.g.*, pictures of accidents)
  - KYC (Know Your Customer) (*e.g.*, fake identity)
  - Impostoring (*e.g.*, impersonating a celebrity)
- **Media creation**
  - Use of deepfakes for special effects
  - Green screens, media processing and composition
  - GAN (Generative Adversarial Network) images
  - Short content bursts
  - UGC (User Generated Content) *e.g.*, TikTok, Triller, Adobe Spark
  - Media tracing, *e.g.*, provenance, content versioning, context
  - Picture and movies production
- **Media modification**
  - Image editing software
  - Movie preservation
  - Film enhancement
  - Restoration of old movies

Based on the above, the following sections provide an overview of illustrative use cases. Both the topics and use cases will be extended in the future based on future feedback and contribution from stakeholders.

A. Misinformation and disinformation

Media usage in breaking news: In his/her coverage, a journalist wants to use images from a social media post depicting police violence during protests. The journalist has to make a fast decision but, naturally, he/she wants to be sure the image in the post is genuine and taken at the mentioned place and time.

Deepfake detection: A news host wants to double check if a video he/she received of the president making questionable claims is genuine and not a deepfake.

Content authenticity checking: An investigative journalist wants to verify if an image depicting past atrocities is actually from that era and place. Such investigations may also pertain to ongoing investigations where human rights staff are receiving media from anonymous sources and need to check whether the depicted atrocities/human rights violations have indeed happened in order to further escalate the investigation.

Content usage tracing: A photographer wants to find out where and how some of the images from his portfolio have been used and check whether they are used in a genuine context.

Academic research: An academic journal reviewer might want to know if an image used as evidence for a successful experiment hasn't been altered and is accurate.

Photographic framing: A journalist received images of the Grand Place in Brussels in the aftermath of the terrorist attacks. Due to the specific framing, the images give a frightening impression of the situation. Therefore, the journalist wants to compare with other images taken at the same place and time but from different perspectives to better evaluate the actual situation.

B. Forgery/media forensics

Insurance fraud: In the context of insurance fraud, an insurer might want to check whether an image used as evidence of an accident has not been manipulated.

Mileage reporting photo: A car insurance company provides a discount program for the customer of limited annual mileage and demands the annual-reporting photo showing the mileage and the time displayed on the front panel of the customer's car. This insurance company might want to check whether the photo reported has not been manipulated.

Photo for cost charge: A series of before and after photos is frequently used for charging repair-costs in modern digital society. In this case, the integrity of a series of photos with the timing information from the origin to the final needs to be authenticated.

Evidence of trial: A prosecutor wants to verify whether a movie recorded by a Closed Circuit TV Security System was really taken at the location and the time claimed.

Media sharing on social media: A media consumer (end user) wants to verify the credibility of a news article shared on a social media and he/she would like to trace who created, modified and published an image included in it. An additional check that the end user would be interested in is to verify if the user account profile picture is authentic or synthetically generated from an AI model (*e.g.*, StyleGAN like that used by `thispersondoesnotexist.com`).

Credibility of AI training image data sets: Online auction service buys a set of training image data from a

stock photo service and wants to check if each image was really taken by a camera instead of being created synthetically.

C. Media creation

Movie special effects: A creative movie production company has created several shots for a movie that are computer generated but almost indistinguishable from real footage. The footage is labeled to allow consumers to identify that the content is computer generated. Since the final movie is a composition of generated and real footage, the entire movie can be labeled frame by frame.

Media transcoding: A photographer develops multiple versions of an image for different purposes. This includes the camera RAW image, rendered JPEG, moderately enhanced image and several versions with varying quality for web preview or print. During each transcoding step, authenticity and IPR information is retained from the parent version to the child version. In addition, authenticity information might be updated to describe modifications inherent to the transcoding process such as loss of quality when transcoding to a lossy format.

Chroma keying or silhouette extraction: Using chroma keying or silhouette extraction, a reporter can be virtually placed in a different location. Labeling the content allows media consumers to identify whether the shots were actually taken at the location or not.

D. Media modification

Image colorization and restoration: A developer has created an algorithm that uses deep learning to colorize grayscale images and enhances the image quality. The output images are labeled to allow consumers to identify that these images have been processed and may not accurately reflect the original colors.

Photo editing: A photographer uses photo editing software (*e.g.*, Photoshop) to edit model pictures for a magazine. The final images are labeled to indicate that they are post-processed. The labels allow to signal how "severe" the changes are to distinguish simple contrast and tone enhancements from changes where content has been added, removed or modified.

V. Requirements

Although still preliminary, based on the identified use cases, a number of JPEG Fake Media requirements have been identified and organized in two main categories, namely, the modification description and the secure linking of modification descriptions and media content. The sections below list the already identified requirements for each identified category.

A. Modification description

- The standard shall provide means to describe how, by who or when the content was created, generated and/or modified.
- The standard shall provide means to describe the type of modification, *e.g.*, no modifications, transcoded, enhanced, restored, colorized, edited, composed, and deep faked.
- The standard shall provide means to describe the purpose of a modification.
- The standard shall provide means to describe (algorithmically or by humans) the probability of a modification.
- The standard shall provide means to describe the region where the media asset was modified.
- The standard shall provide means to attach provenance information to media assets.
- The standard shall provide means to keep track of the history of media asset modifications.
- The standard shall provide means to compress embedded descriptions.
- The standard shall provide means to embed references to externally hosted descriptions.

B. Secure linking of modification descriptions and media content

- The standard shall provide means to restrict access to media asset metadata.
- The standard shall provide means to identify if the media asset has been modified.
- The standard shall provide means to record and protect IPR information and/or provenance information.
- The standard shall provide means to identify the origin of the media asset.
- The standard shall provide means to verify the authenticity of the media asset.

VI. Roadmap for Standardization: Next Steps

The key steps in the definition of a new JPEG standard are illustrated in Fig. 1. The process as depicted, follows a well defined procedure composed of the following steps:

1) Inform and engage stakeholders
2) Collect relevant use cases
3) Assess and organize use cases
4) Define and cluster requirements from use cases
5) Define appropriate performance assessment processes and metrics
6) Issue a Call for Proposals

Engagement with stakeholders is crucial to achieve a good understanding of their specific use cases, requirements and the challenges imposed by the latter. To this end, JPEG already organised a 1st JPEG Fake Media Workshop[3] and more workshops will be organized to further complete the list of use cases with related requirements and challenges.

Next in the process of creating of a standard is the definition and clustering the requirements that the standard should comply to. At this point, as introduced in the previous section, generic requirements have been identified in two categories:

[3] 1st JPEG Fake Media Workshop Proceedings, ISO/IEC JTC1/SC29/WG1, wg1n90026, Online, December 15th, 2020.

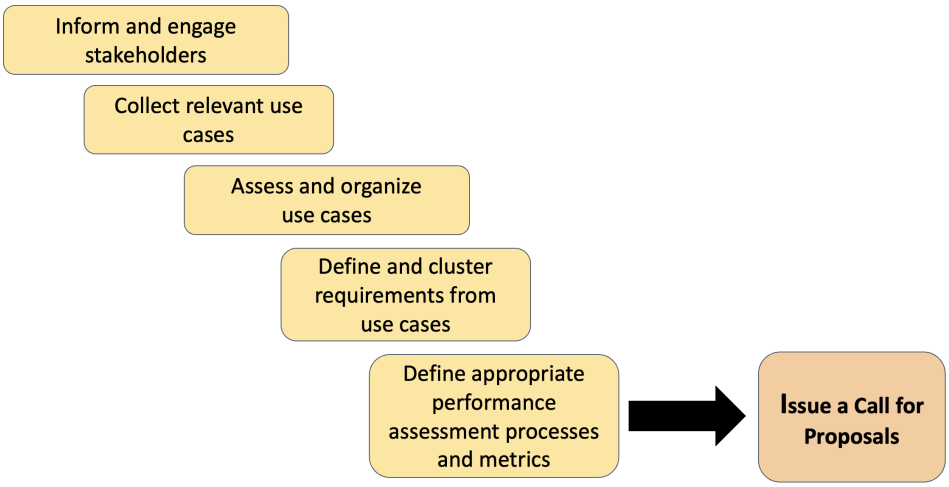

Fig. 1: JPEG standardization process.

modification description and secure linkage of these descriptions to the media content. These requirements will be further refined to identify which requirements can be taken care of by JPEG.

Based on the use cases and the final set of requirements, a *Call for Proposals* will be issued, allowing experts to propose technical solutions that address such requirements. The *Call for Proposals* will include the appropriate assessment processes and metrics that will be used to evaluate the received proposals. After evaluation, the selected technical solutions will be collaboratively improved and completed and might be included in a new standard or might lead to extensions of other relevant standards such as the JPEG Universal Metadata Box Format (ISO/IEC 19566-5) or JPEG Privacy and Security (ISO/IEC 19566-4).

VII. Conclusion

Nowadays, media assets can be modified or even entirely generated in such a way that they are almost indistinguishable from original assets. The existence of modified and generated media have become much more common due to state-of-the-art deep learning techniques that make it much faster and accessible to create synthetic media. While this enables new opportunities for creative usage, intentional or unintentional spread of manipulated media may have serious consequences. Providing consumers the ability to verify the origin and provenance of media assets can alleviate these risks. However, this can only be achieved if security, reliability and interoperability are guaranteed. In this context, this paper intends to introduce the recently launched JPEG initiative to engage with modified media stakeholders to explore usage scenarios with good intention, as well as those with malicious intent. From these usage scenarios, requirements are derived that will eventually lead to the creation of a standard that facilitates secure and reliable annotation of media asset generation and modifications. Interested experts are invited to learn more about the JPEG Fake Media initiative on the JPEG website[4] and to get involved by participating in the discussions[5] and workshops.

Acknowledgment

Several JPEG experts involved in the JPEG Fake Media initiative have contributed to parts of this paper, whose input are hereby acknowledged. The list includes: Nabajeet Barman, Sabrina Caldwell, Spencer Cheng, Symeon Papadopoulos, Leonard Rosenthol, Paweł Korus, Michael W. Steidl, Eduardo A. B. da Silva and Kazuhiko Takabayashi.

References

[1] Dominik Schraml, "Physically based synthetic image generation for machine learning: a review of pertinent literature," in *Photonics and Education in Measurement Science 2019*. International Society for Optics and Photonics, 2019, vol. 11144, p. 111440J.

[2] Miao Wang, Xu-Quan Lyu, Yi-Jun Li, and Fang-Lue Zhang, "VR content creation and exploration with deep learning: A survey," *Computational Visual Media*, vol. 6, no. 1, pp. 3–28, 2020.

[3] Seungjoo Yoo, Hyojin Bahng, Sunghyo Chung, Junsoo Lee, Jaehyuk Chang, and Jaegul Choo, "Coloring with limited data: Few-shot colorization via memory augmented networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11283–11292.

[4] Mohammad Mahdi Johari and Hamid Behroozi, "Context-aware colorization of gray-scale images utilizing a cycle-consistent generative adversarial network architecture," *Neurocomputing*, vol. 407, pp. 94–104, 2020.

[5] David Güera and Edward J Delp, "Deepfake video detection using recurrent neural networks," in *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 2018, pp. 1–6.

[6] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu, "Celeb-df: A large-scale challenging dataset for deepfake forensics," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3207–3216.

[7] Jacquelyn Burkell and Chandell Gosse, "Nothing new here: Emphasizing the social and cultural context of deepfakes," *First Monday*, 2019.

[8] Jessica Ice, "Defamatory political deepfakes and the first amendment," *Case W. Res. L. Rev.*, vol. 70, pp. 417, 2019.

[9] Brendan Sasso, "Virginia journal of criminal law," 2020.

[4] https://jpeg.org/jpegfakemedia

[5] http://listregistration.jpeg.org

[10] Deepayan Bhowmik and Frederik Temmermans, "JPEG White paper: Towards a Standardized Framework for Media Blockchain and Distributed Ledger Technologies," Tech. Rep. WG1N84038, ISO/IEC JTC1/SC29 WG1, 2019.

[11] Justus Thies, Michael Zollhofer, Marc Stamminger, Christian Theobalt, and Matthias Nießner, "Face2face: Real-time face capture and reenactment of RGB videos," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2387–2395.

[12] Justus Thies, Michael Zollhöfer, and Matthias Nießner, "Deferred neural rendering: Image synthesis using neural textures," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1–12, 2019.

[13] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner, "Faceforensics++: Learning to detect manipulated facial images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 1–11.

[14] Shruti Agarwal, Hany Farid, Yuming Gu, Mingming He, Koki Nagano, and Hao Li, "Protecting world leaders against deep fakes.," in *CVPR Workshops*, 2019, pp. 38–45.

[15] Jingjing Zhuge, Jiajun Lin, and Wei An, "Automatic colorization using fully convolutional networks," *Journal of Electronic Imaging*, vol. 27, no. 4, pp. 043025, 2018.

[16] Sanae Boutarfass and Bernard Besserer, "Improving cnn-based colorization of b&w photographs," in *IEEE 4th International Conference on Image Processing, Applications and Systems (IPAS)*, 2020, pp. 96–101.

[17] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao, "Cycleisp: Real image restoration via improved data synthesis," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2696–2705.

[18] Carl Vondrick, Hamed Pirsiavash, and Antonio Torralba, "Generating videos with scene dynamics," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, 2016, pp. 613–621.

[19] Sasan Asadiabadi, Rizwan Sadiq, and Engin Erzin, "Multimodal speech driven facial shape animation using deep neural networks," in *2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 2018, pp. 1508–1512.

[20] Aliaksandr Siarohin, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci, and Nicu Sebe, "Animating arbitrary objects via deep motion transfer," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2377–2386.

[21] Deepali Aneja, Alex Colburn, Gary Faigin, Linda Shapiro, and Barbara Mones, "Modeling stylized character expressions via deep learning," in *Asian conference on computer vision*. Springer, 2016, pp. 136–153.

[22] Amelia Heathman, "IBM Watson creates the first ai-made film trailer—and it's incredibly creepy," https://www.wired.co.uk/article/ibm-watson-ai-film-trailer.

[23] Daniel Holden, Jun Saito, and Taku Komura, "A deep learning framework for character motion synthesis and editing," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1–11, 2016.

[24] Ben Lindbergh, "How they made it: The deeply real deepfakes of 'for all mankind'," March 2021, [Online; posted 5-March-2021].

[25] Leonard Rosenthol, Andy Parsons, Eric Scouten, Jatin Aythora, Bruce MacCormack, Paul England, Marc Levallee, Jonathan Dotan, Sherif Hanna, Hany Farid, and Sam Gregory, "The Content Authenticity Initiative: Setting the Standard for Digital Content Attribution," Tech. Rep., Adobe, CAI, 2020.