# A Media Security Framework Inspired by Emerging Challenges in Fake Media and NFT

Frederik Temmermans[a,b], Deepayan Bhowmik[c], Fernando Pereira[d], and Touradj Ebrahimi[e]

[a]Vrije Universiteit Brussel, Belgium
[b]imec, Belgium
[c]University of Stirling, United Kingdom
[d]Instituto Superior Técnico - Universidade de Lisboa and Instituto de Telecomunicações, Portugal
[e]Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland

## ABSTRACT

Advances in deep neural networks (DNN) and distributed ledger technology (DLT) have shown major influence on media security, authenticity and privacy. Current deepfake techniques can produce near realistic media content which can be used in both good and bad intended use cases. At the same time, DLTs are finding their way in the industry as fair, transparent and reliable means for content distribution. In particular non-fungible tokens (NFTs) are emerging in the digital art market. However, such new developments also introduce new challenges, including the need for robust and reliable metadata, a mechanism to secure the media and associated metadata, means to verify authenticity and interoperability between various stakeholders. This paper identifies emerging challenges in fake media and NFT, and proposes a novel framework to effectively cope with secure media applications allowing for a structured, systematic, and interoperable solution. The framework relies on an architecture that is modular, flexible, extensible, and scalable in the sense that it can be implemented in both lighter as well as more feature-rich and more complex configurations depending on the underlying application, needed features and available resources, while enabling products and services in various ecosystems with desired trust and security capabilities. The framework is inspired by activities and developments within JPEG standardisation related to security, authenticity & privacy.

**Keywords:** Fake media, blockchain, DLT, NFT, media security framework, deepfake, authenticity, provenance, JPEG

## 1. INTRODUCTION

Security, authenticity and privacy concerns have always deserved particular attention in the multimedia landscape. This has been recently reinforced by general awareness due to the rise of social media and incidents reported in traditional media. Historically, most attention has been paid to copyright violation prevention and preservation of privacy. However, recent advances in technologies such as deep neural networks (DNN) and distributed ledger technology (DLT) impose several new challenges. For example, deepfake techniques can produce near realistic media content which can be used in both good and bad intended use cases, e.g. producing new content in the creative sector, restoring and colourising old and noisy media or spreading disinformation. At the same time, DLTs are finding their way in the industry as fair, transparent and reliable means for content distribution. In particular Non-Fungible Tokens (NFTs) are emerging in the digital art market. Challenges include the need for robust and reliable metadata, a mechanism to secure the media content and associated metadata, means to verify authenticity and interoperability between various stakeholders. In addition, a widening image modality landscape with formats such as 360, point clouds, light field and holographic content requires consideration of these challenges from the start.

This paper identifies emerging challenges in fake media and NFT and proposes a media security framework to effectively cope with secure media applications allowing for a structured, systematic, and interoperable solution.

Further author information: send correspondence to frederik.temmermans@vub.be


Frederik Temmermans, Deepayan Bhowmik, Fernando Pereira, and Touradj Ebrahimi "Media security framework inspired by emerging challenges in fake media and NFT", *Proceedings of SPIE 12138,* Optics, Photonics and Digital Technologies for Imaging Applications VII, 121380P (17 May 2022); https://doi.org/10.1117/12.2622223

The framework relies on an architecture that is modular, flexible, extensible, and scalable in the sense that it can be implemented in both lighter as well as more feature-rich and more complex configurations depending on the underlying application, needed features and available resources, while enabling products and services in various ecosystems with desired trust and security capabilities.

Many issues concerning media security, authenticity and privacy have been successfully addressed by standardisation bodies. The proposed framework is inspired by recent activities in JPEG which has a long history of providing not only image coding standards, but also specifications to support imaging ecosystems such as media file formats, metadata handling, and privacy and security features. In particular, two new initiatives, JPEG Fake Media and JPEG NFT, aim to address the above-mentioned emerging challenges.

A holistic and effective solution that can address these challenges is currently missing, which requires flexible integration of multiple functionalities as desired by various stakeholders in the media consumption chain. Inspired by the above mentioned JPEG activities, the proposed framework aims to fill this gap. The remaining sections of this paper are organised as follows: Section 2 describes the status quo related to existing JPEG specifications and explorations to support security, authenticity and privacy enabled workflows. Thereafter, Section 3 discusses fake media and NFT specific emerging challenges. Section 4 introduces the proposed media security framework that can comprehensively accommodate existing challenges in media security, such as ownership and privacy protection, as well as emerging challenges in fake media and NFT. Finally, Section 5 wraps up and draws conclusions.

## 2. STATUS QUO IN JPEG

Many standardization and private industry initiatives focus on security, authenticity and/or privacy in multimedia. Some examples include MPEG-21,[1] the Content Authenticity Initiative (CAI),[2] the Coalition for Content Provenance and Authenticity (C2PA),[3] the Digital Watermarking Alliance (DWA),[4] OpenCA[5] and W3C PROV.[6] Since the framework proposed in this paper is inspired by the JPEG initiatives on Fake Media and NFT, this section focuses on the background of JPEG that led to these initiatives.

JPEG has a long-lasting history with providing media specifications to support imaging ecosystems. While the JPEG 1 format has been around for more than 28 years, it is still the most dominant image file format in the consumer market. Meanwhile, JPEG 2000 plays a pertinent role in professional markets such as medical imaging, digital cinema and geographic information systems. More recently, JPEG XS has been designed for applications requiring transparent quality, low complexity and low latency while JPEG XL is a promising successor to JPEG 1 with improved efficiency and feature set, but also smoothly integrating with JPEG 1 based workflows. While these formats mostly focus on traditional media, JPEG Pleno focuses on emerging modalities such as point clouds, light field and holographic media.

In addition to image coding and representation solutions, JPEG also supports imaging ecosystems by providing specifications for media file formats, metadata handling, and privacy and security features. Related to the latter, JPSEC provides a framework, concepts and methodology for securing JPEG 2000 codestreams.[7] JPSEC provides three types of security services for JPEG 2000 images: confidentiality, integrity, and authentication. JPSEC leverages security methods that are established and proven as secure, and specifies how these security methods are applied to JPEG 2000 images in a media-aware way.[8]

While JPSEC leverages comprehensive security functionalities, the standard is specific to JPEG 2000 images. The aim of JPEG Systems (ISO/IEC 19566) is to provide a metadata and extensions framework that can be used with any of the JPEG coding formats, from the original JPEG 1 format, over JPEG 2000, to the latest formats such as JPEG XS and JPEG XL. The JPEG Universal Metadata Box Format or JUMBF is the foundation on which metadata-based extensions are built within JPEG Systems.[9] JUMBF provides a scheme to define containers to embed metadata in JPEG images in a universal way.[10] This metadata can be textual, but also image content or any type of binary data. JUMBF also provides additional functionalities to facilitate linkage between metadata and image data. This includes signalling the type of embedded content, the creation of references between related contents, integrity checking via checksums and a URL syntax to formalize references and requests to embedded metadata.

JPEG Privacy and Security [9] is a JUMBF based extension that focuses on features related to protection and authenticity. For protection, the standard supports tools to protect parts of images and/or associated metadata while retaining backward and forward compatibility. In practice, this means that a legacy decoder won't be able to decode the protected information, but it will be able to process and render the file as a normal image. Access control can be managed independently for various regions in the image or metadata instances. For authenticity, the standard focuses on the use of signatures or hashes to check integrity. These can apply to specific image regions or to its associated metadata.[11]

In January 2018, JPEG started a new initiative to explore standardization needs related to media blockchains.[12] This ultimately led to another initiative called JPEG Fake Media. The scope of JPEG Fake Media is to produce a standard that can facilitate secure and reliable annotation of media asset creation as well as its modifications, while supporting usage scenarios that are either in good faith or with malicious intent.[13] Finally, JPEG NFT[14] started in April 2021 to explore standardization needs in NFT.

## 3. EMERGING CHALLENGES

Recent emergence of new technologies such as the use of deep learning in imaging has created many opportunities in the creative industries, from content creation to media restoration. Similarly, blockchain technology provides new opportunities to deliver a trusted media consumption chain. Tokenization (e.g. NFT) of digital art for digital asset transactions has made major headlines due to its enormous monetary value. However, these advances also introduce new challanges related to security, authenticity and privacy. This section discusses some of these emerging challenges, in particular, for fake media and NFT.

### 3.1 Fake Media

Multimedia forensics has become even more significant due to the recent rise in both ill intended use cases, such as spreading of fake news through doctored media, as well as use cases in the creative industries, where adoption of AI has shown enormous promises. In practice, deep learning has influenced new applications due to its end to end computational pipeline and higher accuracy when compared to traditional methods. Wide availability of software, e.g. Face2Face,[15] NeuralTextures[16] and FaceSwap [*] enables creation of deepfakes with near realistic contents, almost indistinguishable from captured content to the human eye. These advances assist in the spread of fake news and therefore pose a new set of challenges. The scale of the problem prompted large tech companies like Google to a create deepfake dataset[17] with over 1.8 million manipulated images for open research and multiple governmental bodies around the world currently discuss policies for law enforcement. However, these techniques are also increasingly used in the creative industry. Applications include colourisation of archived media[18,19] or black and white photographs,[20] content creation,[21,22] speech driven facial shape animation,[23] object animation,[24] style transfer[25] and motion synthesis.[26]

The above challenges are addressed by JPEG in the so-called *JPEG Fake Media* initiative that has identified relevant requirements,[13] instrumental for the design of the media security framework proposed in Section 4. These requirements are organized in three main categories: i) requirements of media creation and modification descriptions, which target achieving a transparent and interoperable media production and consumption framework; ii) requirements of metadata embedding and referencing, acknowledging that media asset content and media asset metadata integration plays a key role in the description of the media assets creation and modifications; and iii) requirements related to authenticity, integrity, and trust model under the understanding that an authentic media asset should be verifiable, i.e. able to be checked, and/or trustworthy, i.e. able to be relied on as truthful.

### 3.2 Non-Fungible-Tokens (NFT)

With the advancement of blockchain technology, tokenization of digital assets became popular and increasingly adopted in various applications, e.g. in IoT,[27] real estate[28] or digital art.[29] Unlike physical art where each creation is unique, digital art often suffers from its distinctiveness as digital copy allows to make exact replicas. Recent advances in blockchain technology offer *tokenization* of digital arts[30] where digital assets can be made

[*] github.com/MarekKowalski/FaceSwap/

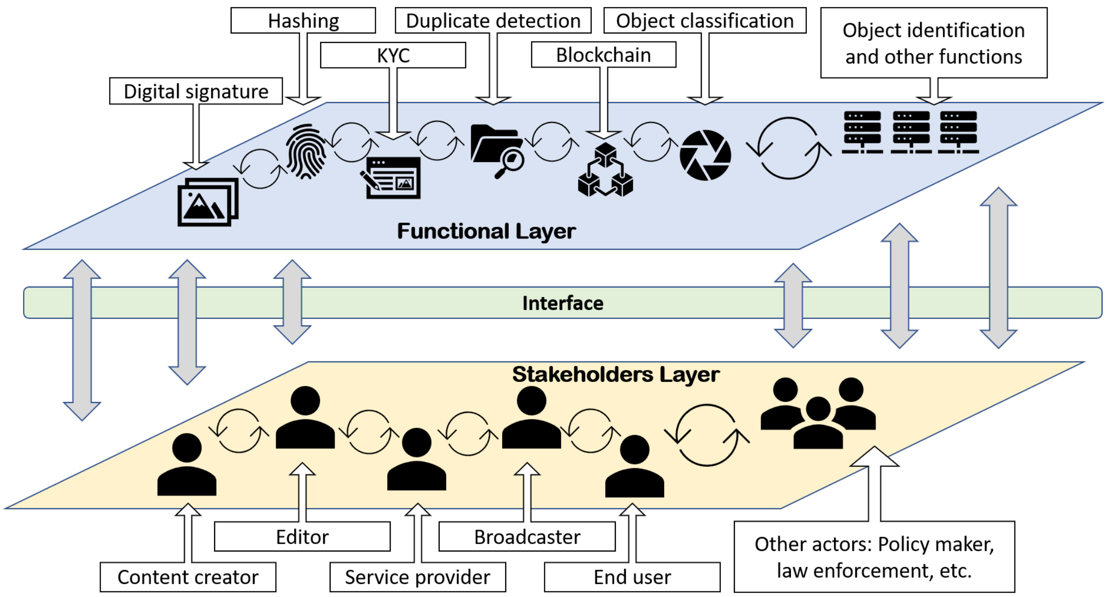


Figure 1. Proposed media security framework.

unique, and consequently bring the concept of scarcity into the digital world through NFT. The latter has made recent headlines, e.g. a digital collage, '*Everydays — The First 5000 Days*', by Peeble which was sold at an auction for $69M. While such advances are welcomed by the creative industry, they also create new challenges, including the need for a standard metadata format for blockchain entries, unique asset signature creation, security, trust, interoperability and legally binding framework, to name a few. Although *tokenized art* is still in its infancy, high demand and high value auction prices indicate the need for a standard framework that can address the above issues. To this end, JPEG has launched the JPEG NFT initiative.[14]

## 4. PROPOSAL FOR A COMPREHENSIVE FRAMEWORK

In this section we propose a novel framework to effectively cope with secure media applications based on both JPEG and non-JPEG compliant standards, allowing for a structured, systematic, and interoperable solution. The proposed framework relies on an architecture that is modular, flexible, extensible, and scalable in the sense that it can be implemented in both lighter as well as more feature-rich and more complex configurations depending on the underlying application, as well as for needed features and available resources, while enabling products and services in various ecosystems with desired trust and security capabilities. The proposed framework is in part similar to the OSI architecture that is widely used in telecommunication systems.[31] Various improvements have been brought to target a more efficient and trusted as well as secure media framework, for applications that involve media-centric solutions. The media security framework is structured in two layers, namely, the stakeholders layer and the functional layer, each composed of a number of key elements. The two layers are the two sides of the same coin and provide complementary perspectives of the ecosystem.

### 4.1 Stakeholders Layer

The **stakeholders layer** defines who are the **actors** in the ecosystem which interact directly. Actors represent all entities which play a role in the ecosystem around media assets. Examples include the creator of a media asset, a device, a product, a service and generally any tools or solutions used to capture or to synthetize the media, the end-user who consumes the content which could be either a human or a machine, the editor of the media asset, the infrastructure or service provider in charge of storage, delivery, broadcasting or distribution of the media asset, the solution in charge of analyzing the media asset for any purpose, service provider in charge of search, annotation and management of the media asset, etc. Actor here is used in the largest possible definition of the term and includes policy makers, law enforcement, governmental and non-governmental agencies that directly or indirectly either interact with the media asset or influence its ecosystem.

## 4.2 Functional Layer

The **functional layer** is an architecture made of **functions** and their **interactions**. Functions process and exchange information, including media assets. Each function is defined by a description of *what it does* and not *how it is implemented.* Examples of functions include simple and generic tools such as digital signature, hash, encryption, as well as more complex solutions for media assets such as (near) duplicate detection, KYC (Know Your Customer), object classification and object identification tools. Interactions are described by the nature of the information being exchanged and the syntax used. This includes the representation used for the media asset but also other types of metadata structure such as JUMBF.

## 4.3 Other Key Components

The functional layer defines which tools and solutions are used when actors interact. The stakeholders layer defines which actors interact in a specific use case, and describes which tools and solutions in the functional layer are employed and which media assets as well as other types of information are exchanged. The latter is defined by means of interfaces, while the former is defined by means of interactions. Below we define in more details some of the key components of the proposed framework.

- **Media assets:** This refers to any digital asset including images, videos, audio or text. In the context of this paper, we mainly focus on visual modality in form of image and video, however, other media types are not excluded and can be part of the proposed framework.
- **Interfaces:** This refers to the structure and the syntax of the information that is exchanged between layers as opposed to interaction within a layer.
- **Interactions:** This refers to interactions within a layer, such as interactions between functions in the *functional layer* or between actors in the *stakeholders layer*. Interactions can be one-to-one, one-to-many or many-to-many.
- **Configurations:** Although the proposed framework as defined here allows for the widest and most flexible types of interactions between any of its actors and functions, for practical reasons, such as guaranteeing interoperability or imposing upper bounds on required resources (e.g. complexity), specific configurations can be defined as part of the framework using a standard syntax as a result of capabilities offered for a specific ecosystem, similar to the notion of profiles and levels in JPEG standards, but in an extended sense where also other architectural constraints can be imposed.

## 4.4 Key Novelties and Benefits

The proposed framework addresses several novelties and advantages that are essential to efficiently deal with security, authenticity and privacy of digital assets. The following lists the most important:

- The proposed framework can cope with any media asset representations, notably compliant or not with any coding standards, their associated metadata and the information exchanged between functions.
- Through its two layers, the proposed framework provides a modular approach to define interactions between actors in the ecosystem, which tools and solutions are used by each actor, and the information and media assets exchanged, in a clear and explicit way, informing what is required to ensure interoperability.
- The proposed framework is versatile and can cope with use cases that have various constraints in terms of tools/solutions and available resources, ranging from lightweight and low-complexity to more feature-rich configurations.
- The combination of functional description as opposed to the exact implementation as defined in the framework, provides flexibility and extensibility, in the sense that additional actors and tools/solutions can be added while ensuring interoperability.

- The proposed framework facilitates the integration of products and services seamlessly, not only enabling interoperability for specific applications, but also across applications and ecosystems that require trust and security.

In this paper, the needs and benefits above have been inspired by ecosystems relying on JPEG standards where not only legacy security and trust requirement (copyright, integrity verification and privacy protection) but also new and emerging requirements can be addressed, such as those arising by fake media and NFT; however, other ecosystems can also benefit from this framework as it is fundamentally standard agnostic.

## 5. CONCLUSION

Emerging technologies such as deep fakes and NFTs have a significant impact on media creation, content editing, distribution and consumption. As a consequence, these new technologies introduce a new set of challenges related to security, authenticity and privacy. Inspired by JPEG initiatives on fake media and NFT that aim to address these challenges, this paper proposed a modular and scalable framework that accommodates various secure media applications in a structured, systematic and interoperable manner. The framework envisages a two-layer approach consisting of a stakeholders layer, defining the actors in the ecosystem, and a functional layer which is an architecture made of functions and their interactions. The framework facilitates seamless integration of many media products and services in an interoperable and secure manner.

## ACKNOWLEDGMENTS

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, Janos Farkas, Paweł Korus, Neal Krawetz, Symeon Papadopoulos, Leonard Rosenthol, Michael W. Steidl, Eduardo A. B. da Silva and Kazuhiko Takabayashi.

## REFERENCES

[1] Burnett, I., Van de Walle, R., Hill, K., Bormans, J., and Pereira, F., "MPEG-21: goals and achievements," *IEEE MultiMedia* **10**(4), 60–70 (2003).

[2] Rosenthol, L., Parsons, A., Scouten, E., Aythora, J., MacCormack, B., England, P., Levallee, M., Dotan, J., Hanna, S., Farid, H., and Gregory, S., "The Content Authenticity Initiative: Setting the Standard for Digital Content Attribution," tech. rep., Adobe, CAI (2020).

[3] "Coalition for Content Provenance and Authenticity (C2PA)." https://c2pa.org. Accessed: 2021-06-25.

[4] "Digital Watermarking Alliance (DAW)." https://digitalwatermarkingalliance.org. Accessed: 2021-06-25.

[5] Lee, J., Chen, A., Laurent, P., Lehrer, M., and Siedle, D., "Openca: Conditional access system in mediaflo™," in [*2008 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*], 1–6, IEEE (2008).

[6] "W3C PROV." https://www.w3.org/TR/prov-overview/. Accessed: 2021-06-25.

[7] ISO/IEC, "JPEG 2000 image coding system: Secure JPEG 2000," Tech. Rep. 15444-8, ISO/IEC JTC1/SC29 WG1 (2007).

[8] Schelkens, P., Skodras, A., and Ebrahimi, T., [*The JPEG 2000 Suite*], Wiley Publishing (2009).

[9] ISO/IEC, "JPEG systems — Part 4: Privacy and Security," Tech. Rep. 19566-4, ISO/IEC JTC1/SC29 WG1 (2020).

[10] Temmermans, F., Kuzma, A., Choi, S., and Schelkens, P., "Adopting the JPEG systems layer to create interoperable imaging ecosystems," in [*Optics, Photonics and Digital Technologies for Imaging Applications VI*], Schelkens, P. and Kozacki, T., eds., **11353**, 177 – 183, International Society for Optics and Photonics, SPIE (2020).

[11] Temmermans, F., Bhowmik, D., Pereira, F., Ebrahimi, T., and Schelkens, P., "Exploration of media block chain technologies for JPEG privacy and security," in [*Optics, Photonics and Digital Technologies for Imaging Applications VI*], Schelkens, P. and Kozacki, T., eds., **11353**, 177 – 185, International Society for Optics and Photonics, SPIE (2020).

[12] JPEG, "JPEG White paper: Towards a Standardized Framework for Media Blockchain and Distributed Ledger Technologies," Tech. Rep. WG1N84038, ISO/IEC JTC1/SC29 WG1 (2019).

[13] JPEG, "JPEG Fake Media: Context, Use Cases and Requirements," Tech. Rep. WG1N91018, ISO/IEC JTC1/SC29 WG1 (2021).

[14] "1st JPEG NFT Workshop Announcement." https://jpeg.org/items/20210616_1st_jpeg_nft_workshop_announcement.html. Accessed: 2021-06-25.

[15] Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., and Nießner, M., "Face2face: Real-time face capture and reenactment of RGB videos," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 2387–2395 (2016).

[16] Thies, J., Zollhöfer, M., and Nießner, M., "Deferred neural rendering: Image synthesis using neural textures," *ACM Transactions on Graphics (TOG)* **38**(4), 1–12 (2019).

[17] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., and Nießner, M., "Faceforensics++: Learning to detect manipulated facial images," in [*Proceedings of the IEEE/CVF International Conference on Computer Vision*], 1–11 (2019).

[18] Johari, M. M. and Behroozi, H., "Context-aware colorization of gray-scale images utilizing a cycle-consistent generative adversarial network architecture," *Neurocomputing* **407**, 94–104 (2020).

[19] Zhuge, J., Lin, J., and An, W., "Automatic colorization using fully convolutional networks," *Journal of Electronic Imaging* **27**(4), 043025 (2018).

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

[21] Wang, M., Lyu, X.-Q., Li, Y.-J., and Zhang, F.-L., "VR content creation and exploration with deep learning: A survey," *Computational Visual Media* **6**(1), 3–28 (2020).

[22] Vondrick, C., Pirsiavash, H., and Torralba, A., "Generating videos with scene dynamics," in [*Proceedings of the 30th International Conference on Neural Information Processing Systems*], 613–621 (2016).

[23] Asadiabadi, S., Sadiq, R., and Erzin, E., "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)*], 1508–1512, IEEE (2018).

[24] Siarohin, A., Lathuilière, S., Tulyakov, S., Ricci, E., and Sebe, N., "Animating arbitrary objects via deep motion transfer," in [*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*], 2377–2386 (2019).

[25] Aneja, D., Colburn, A., Faigin, G., Shapiro, L., and Mones, B., "Modeling stylized character expressions via deep learning," in [*Asian conference on computer vision*], 136–153, Springer (2016).

[26] Holden, D., Saito, J., and Komura, T., "A deep learning framework for character motion synthesis and editing," *ACM Transactions on Graphics (TOG)* **35**(4), 1–11 (2016).

[27] Arcenegui, J., Arjona, R., Román, R., and Baturone, I., "Secure combination of iot and blockchain by physically binding iot devices to smart non-fungible tokens using pufs," *Sensors* **21**(9), 3119 (2021).

[28] Dowling, M., "Fertile land: Pricing non-fungible tokens," *Finance Research Letters* , 102096 (2021).

[29] Chow, A. R., "Nfts are shaking up the art world—but they could change so much more," (March 2021). [Online; posted 22-June-2021].

[30] Trautman, L. J., "Virtual art and non-fungible tokens," *Available at SSRN 3814087* (2021).

[31] Zimmermann, H., "OSI Reference Model - The ISO Model of Architecture for Open Systems Interconnection," *IEEE Transactions on Communications* **28**(4), 425–432 (1980).