Blockchain:
la solució actual vers el vídeo hipertrucat

Treball Final de Grau: Revisió Bibliogràfica

Estudiant: Jordi Muñiz
Pla d'estudis: Grau en Informació i Documentació

Tutor del Treball Final de Grau: Dr. Alexandre López-Borrull

Data de Lliurament: 05/06/2021


BY NC SA

RESUM

Introducció: Es mira de respondre si la tecnologia disruptiva anomenada *blockchain* (cadena de blocs), pot aportar solucions enfront el fenomen *deepfake* (vídeo hipertrucat), per solucionar els problemes d'autenticació de les imatges en un context de desinformació.

Metodologia: Es realitza una revisió bibliogràfica a repositoris, biblioteques digitals i base de dades de ciències físiques, d'informàtica i enginyeria i d'altres de recerca multidisciplinària. Es trien 16 articles significatius per a l'anàlisi.

Resultats: La comunitat científica i acadèmica planteja el *deepfake* com amenaça, pel seu impacte social, polític i econòmic. Les característiques pròpies de la cadena de blocs, com són la traçabilitat, el seguiment, la transparència i la confiança; la fan òptima per a autenticar i frenar el fenomen *deepfake*.

Conclusió: Les xarxes de descentralització per a l'emmagatzematge i compartiment d'arxius distribuït, les DLT, i en concret la cadena de blocs - amb les característiques tècniques i metodològiques, que li són pròpies com són l'ús de contractes intel·ligents, les proves d'autenticitat, les empremtes digitals criptogràfiques i d'altres - ha de combinar-se amb sistemes de detecció amb intel·ligència artificial, amb la tecnologia de les marques d'aigua i amb la complicitat i compromís de les grans empreses tecnològiques i plataformes de xarxes socials, en l'autentificació i preservació del contingut digital pristí i en la detecció de la seva manipulació.

Paraules clau: *vídeo hipertrucat, deep fake, deepfake video, blockchain, cadena de blocs, hash, marques d'aigua, intel·ligència artificial, digital forensics*

ABSTRACT

Introduction: Blockchain disruptive technology maybe provides solutions to the phenomenon of deepfake and try to solve the problems of image authentication in a post-truth era.

Methodology: A bibliographic review search in repositories, digital libraries and databases of physical sciences, computer science and engineering and others of multidisciplinary research. This study is selected sixteen significant articles for analysis.

Results: The scientific and academic community poses deepfake's phenomenon as a menace because of its social, political and economic impact. The characteristics of the blockchain, such as traceability, tracking, transparency and trust, make it optimal to authenticate and curb the deepfake phenomenon.

Conclusion: Goals with distributed file systems, peer-to-peer networks for sharing and storing, DLTs and blockchain's framework seems satisfactory. Blockchain technical and methodological characteristics, which are its own such as the use of smart contracts, proof-of-authority, hash values and other parameters, must be combined with artificial intelligence detection systems, watermark technology and complicity and commitment of large tech companies and social media platforms, in the authentication and preservation of digital pristine content and detection of its tampering.

Keywords: *deep fake, deepfake video, blockchain, hash, watermarks, artificial intelligence, digital forensics*

Abreviatures

AI: Artificial Intelligence

CNN: Convolutional Neural Network

DLT: Distributed Ledger Technologies

DoS: Denial of Service

GAN: Generative Adversarial Networks

HF: Hyperledger Fabric

IPFS: Inter-Planetary File System

IoT: Internet of Things

LOCKSS: Lots of Copies Keep Stuff Safe

LSTM: Long Short-Term Memory

ML: Machine Learning

P2P: Peer-to-Peer

PoA: Proof-of-Authenticity

PoA: Proof-of-Authority

PoW: Proof-of-Work

PoS: Proof-of-Stake

RNN: Recurrent Neuronal Network

SIFT: Scale-Invariant Feature Transform

SM: Smart Contract

SSI: Self-Sovereign Identity

Agraïments

Voldria tenir unes paraules d'agraïment al Dr. Alexandre López-Borrull pel tutoratge brindat en aquest treball i per ajudar-me a definir les línies del desenvolupament d'aquest.

A la Dra. Montserrat Garcia Alsina, per donar-me la oportunitat de fer el pràcticum universitari a un centre de documentació audiovisual.

També agrair a Aida Camps i a la resta de l'equip de la Biblioteca per a l'Aprenentatge de la UOC, que m'ha donat suport durant aquests darrers mesos.

I al meus companys de promoció de tots aquest anys, Susana Castellana i Ricard Guasch.

Gràcia, juny de 2021

*"*Our works in stone, in paint, in print, are spared, some of them, for a few decades or a millennium or two, but everything must finally fall in war, or wear away into the ultimate and universal ash; the triumphs, the frauds, the treasures and the fakes. A fact of life: we are going to die. '*Be of good heart*,' cry the dead artists out of the living past. Our songs will all be silenced, but what of it? '*Go on singing*'…Maybe a man's name doesn't matter all that much*"*.

Orson Welles, 1973, *'F for Fake'*

Índex

1. INTRODUCCIÓ

Durant l'estada al Centre de Documentació de RTVE, en la realització del pràcticum universitari, van sorgir inquietuds i dubtes de com es podrien protegir o certificar l'autenticitat dins d'un repositori audiovisual, no solament des d'un punt de vista de vigència documental sinó amb les implicacions a llarg termini que això podria comportar.

Quan es treballa en un repositori audiovisual s'accepta l'apriorisme de l'autenticitat del document original - o pristí - que en aquest cas esdevé el suport de gravació que l'ENG (*electronic news gathering*, la recopilació electrònica de notícies d'àudio i vídeo), ja sigui el periodista, el redactor o el cameràman, autentifiquen i donen fe als documentalistes que ingesten, minuten i descriuen, ja "que les metadades o informació addicional s'han d'introduir en la mesura del que sigui possible de la font" (Agirreazaldegi, 2011). I això es pot posar en risc.

S'assenyala que en un futur l'AI (intel·ligència artificial) automatitzarà, tots aquests processos de manera necessària, en un context de massificació d'arxius visuals. I així ho mostren informes de govern i consultores privades de les tecnologies de la informació, com Gartner, que visualitzen l'impacte i la "tendència cada cop més intensiva" (Marcet, 2020) a diversos sectors tant de l'àmbit de la indústria pública i privada i de la recerca, que indiquen una reducció dràstica en les "tasques manuals associades a la gestió de dades" en els propers anys (Serra, 2020).

Un dels exemples de recerca, podria ser la *Venice Time Machine*, projecte liderat per Frederic Kaplan – especialista, entre d'altres, en la cartografia de la història urbana mitjançant la semàntica - que com si fos d'un projecte d'història extractiva, i amb la utilització d'un gran volum de dades de deu segles de la ciutat de Venècia, vol permetre un "model multidimensional" de la ciutat per ser estudiat i aplicat en camps com "l'educació i les ciències històriques" (Kaplan, 2015). Un ús eficient i creixent, en aquest àrea, del Big Data, permetrà "mirades sobre molts períodes històrics" i "molts passatges de la historia afloraran amb molta més complexitat" (Marcet, 2018). Altres projectes com la JFK Files de

Microsoft es parla de cerca cognitiva, on amb l'emmagatzematge al núvol *Azure Search*, i l'ús d'intel·ligència artificial, proporcionant eines d'indexació, consulta, i cerca complexa amb un gran flux de dades, on s'interpreten documents, fotografies, i altres tipus de material per extreure i comprendre informació (Microsoft, 2018).

Per tant, la protecció davant la intrusió i manipulació en el nostre imaginari col·lectiu, que son els repositoris audiovisuals i de banc d'imatges, es fa necessària.

En el context actual de creixent consum de necessitats informatives a través de les xarxes socials i en línia (Nielsen *et al.,* 2020), de notícies falses i de post-veritat, de precarietat en la feina del periodista/creador de contingut, fa que la il·lustració de les notícies amb imatges, i la redacció de les mateixes, es faci ràpid i a barrisc (Ufarte, 2012), amb d'altres variables que queden fóra de la intenció d'aquest treball. I han aparegut amb força d'altres factors que es consideren claus i que centraran el present TFG.

Un primer factor són els avenços que han esdevingut en el camp de l'AI (intel·ligència artificial), del *deep learning*, i més concretament en el camp del *machine learning*, l'aparició de les GANs (*generative adversarial networks*), les xarxes generatives neuronals, que han permès i facilitat entrenar models fent-n'hi ús d'una gran quantitat de dades d'imatges i vídeos.

Això ha obert les portes a la creixent utilització i difusió - gràcies també a la comoditat i simplicitat d'ús de les aplicacions informàtiques - de l'anomenat fenomen *deepfake* de vídeo i de la imatge - resultant de l'hipertrucatge, "aquell conté situacions fictícies amb una gran aparença de realitat, utilitzant sovint amb l'objectiu de desinformar o difamar". Una definició que ens aporta TERMCAT i que amb el pas dels dies, esdevé curta. Per què la 'gran aparença de realitat' està sent ja indistingible per a qualsevol ull humà i pels primerencs sistemes de detecció que es van proposar durant la darrera dècada.

I per fer-hi front un altre factor important en aquest context, que centrarà la cerca d'aquest treball, l'anomenada nova tecnologia disruptiva, la cadena de blocs o *blockchain*, una de les tecnologies emergents que han aparegut en els darrers anys.

En la literatura científica i en concret amb les TIC, s'esmenta habitualment al científic nord-americà Roy Charles Amara, que dona nom a la llei que ell mateix exposava en la que "tendim a sobrevalorar l'efecte d'una tecnologia a curt termini i subestimar els seu efecte a llarg termini" (Amara's Law), i s'haurà d'esclarir si és el cas d'ambdós factors.

1.1. Objectiu del treball

La pregunta d'investigació és:
Què pot aportar la metodologia *blockchain* (cadena de blocs) per solucionar els problemes d'autentificació de les imatges vers la tendència actual de deepfake (vídeo hipertrucat)?

L'objectiu doncs és:
Identificar els avantatges que comporta la tecnologia *blockchain*, per solucionar els problemes d'autentificació de les imatges.

2. MARC TEÒRIC

2.1. La cadena de blocs

Hi ha consens general en què la *blockhain* és una tecnologia disruptiva, entesa com aquella innovació que provoca una interrupció amb l'estat actual de les coses, i acaba per substituir una tecnologia anterior (Ab Rahman, 2017). Alguns autors, ja la consideren "la tecnologia emergent més rellevant des de l'aparició d'internet" (Dans, 2017).

Aquesta tecnologia, té una data de naixement, que és l'any 2008, i que alguns situen amb el paper "Bitcoin: A Peer-to-Peer Electronic Cash System" (Nakamoto, 2008) de Satoshi Nakamoto, autor de biografia desconeguda (García-Morales, 2018). En aquest article es va definir "una estructura distribuïda entre iguals, que es podria utilitzar per resoldre el problema de mantenir l'ordre de les transaccions i evitar el problema de la doble de despesa" (Casino *et al*., 2019) en anglès anomenat *double spending*.

Si ens traslladem al món real (per exemple quan realitzem una operació compra-venda entre un venedor i un comprador) podrem explicar millor en què consisteix el *double spending* o problema de la doble despesa, ja que no és possible fer diferents pagaments amb la mateixa "moneda o bitllet a dues persones a la vegada" però en el "en el món digital és molt fàcil copiar registres electrònics per la qual cosa és molt complicat evitar-ho", amb qual cosa es necessitava, amb els pagament electrònics, confiar en una tercera entitat[1], com una entitat bancària (Sarias, 2020). Amb la tecnologia *blockchain* ja no caldrà.

Malgrat això, els seus orígens i alguns del seus fonaments podríem trobar-los en diverses fonts, destacant-ne quatre:

- La publicació del *paper* (article) "New directions in cryptography", que recollia les idees de Whitfield Diffie i Martin Hellman, l'any 1976, d'una criptografia de clau pública (*cryptography public key*), mitjançant un

[1] "la criptologia substitueix els intermediaris de tercers com a custòdia de la confiança, ja que tots els participants de la cadena de blocs executen algoritmes complexos per certificar la integritat del conjunt" (Grech i Camilleri, 2017).

algorisme unidireccional (Moore i Rid, 2016), que permetria la creació per parells de claus (*cryptography two-key cryptography*), ja que tant l'emissor com el receptor podien tenir claus públiques i claus privades, aquestes darreres desconegudes pels altres. La criptografia de clau pública també és anomenada com *asymmetric key* (Hongling i Di, 2019).

- La publicació del *paper* "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms" de David Chaum, l'any 1981 on exposava una criptografia de clau pública per a un sistema de correu electrònic, i que permetria l'anonimat de qui es comunica i el contingut de la seva comunicació.

- La publicació del *paper* "Formalizing and Securing Relationships on Public Networks" de Nick Szabo, l'any 1997, on mitjançant l'exemple de la màquina expenedora, concebeix el concepte dels *smart contract* (en català anomenats contractes intel·ligents), que serien programes informàtics que estableixen els termes de transacció on s'emmagatzemen i que formen una part consubstancial de la cadena de blocs. Els que participin d'aquest acord, accepten executar aquest codi, i això haurà de garantir que, la transacció i les condicions de la seva interacció, "produirà els mateixos resultats per a tots els participants" (Solomon, 2019). Alguns autors consideren que Szabo introdueix el concepte al 1994 (Tapscott i Tapscott, 2017).

- La publicació del *paper* "B-Money-an Anonymous, Distributed Electronic Cash System" de Wei Dai, l'any 1998, que proposa, emprant la criptografia com a mesura de control, la creació d'un nou tipus de moneda descentralitzada.

Tots aquests avenços, entre molts d'altres, va donar llavors a filosofies i moviments filosòfics de ciberactivisme com són el *cypherpunk* i el criptoanarquisme (Marqués-Pascual, 2020) - a favor de la privacitat en front el poder i del seu control i vigilància - i que pren expressió mitjançant el *bitcoin*, la criptomoneda que utilitza la tecnologia *blockchain*, el mateix any que el text de

Sakamoto i de la fallida del banc *Lehmann Brothers* i la seva posterior crisis, al 2008. Que amb molta probabilitat, no deuria ser una casualitat.

Com assenyala Ruipérez García (2019), la barreja de tots aquest factors - de corrent filosòfica, amb transversalitat de coneixements, processos de treball, conceptes matemàtics i criptogràfics, que és confon com a sinònim de criptomoneda – fa un *totum revolutum* que en dificulta la comprensió i fa que un s'apropi a la cadena de blocs amb profusió i confusió de conceptes[2]. S'ofereixen diverses expressions de concepte bàsic i altres elements que la componen:

- Definició de cadena de blocs que proposa la consultora Gartner (Panetta, 2020):

  "És un tipus de llibre major distribuït[3], una llista ordenada cronològicament de registres transaccionals irrevocables i signats criptogràficament compartits per tots els participants d'una xarxa."

- Definició de cadena de blocs que proposa García-Morales (2018) citant a Pullicino (2017):

  "una tecnologia de base de dades descentralitzada i distribuïda[4] que permet mantenir un registre creixent de transaccions mitjançant encriptació i altres activitats, i verificant la seva permanència i historial de canvis"

- Definició de cadena de blocs que proposa (Zamorano, 2018)

[2] S'ha arribat a parlar de la cadena de blocs com la màquina de confiança (*Trust Machine*, Alex Winter, 2018), perquè permet que "les persones que no tenen una confiança especial les unes amb les altres col·laborin sense haver de passar per una autoritat central neutral [i per tant] és una màquina per crear confiança" (The Economist, 2015).

[3] En comptabilitat, és el llibre major (*ledger* en anglès) o "llibre principal en què s'anoten les dades corresponents al govern econòmic d'una empresa" (DIEC). En el nostre cas és on queden registrades tot tipus transaccions o "qualsevol altra qüestió que es pugui descriure en format digital" (Grech i Camilleri, 2017).

[4] També es coneix per DLT, que és l'acrònim de *Distributed Ledger Technologies*, el que es podria traduir com a tecnologies distribuïdes de llibre o registre comptable, que serveix per gestionar dades i actius digitals. La cadena de blocs, pertany a aquesta família de SLT, o *Shared Ledger Technologies* (García-Morales, 2018). Promou la idea d'una xarxa "descentralitzada" enfront el mecanisme convencional de centralització, i no hi ha la necessitat que una autoritat central en reguli el control (Frankenfield i Rasure, 2021).

> "protocol d'operacions d'intercanvi, entre parells[5], que es produeix, es gestiona i es comprova de forma descentralitzada, automatitzada, compartida i segura" entre nombrosos ordinadors, sense que hagin de passar per un tercer[6].

Donats aquests conceptes preliminars es troba adient oferir una síntesi explicativa sobre el procés metodològic de la cadena de blocs. De tota la bibliografia consultada destaca la següent:

> **"**El nom *blockchain* prové de el fet que totes les dades s'emmagatzemen en blocs, i cada bloc està connectat al bloc anterior, formant una estructura en forma de cadena [*chain*]. Només es pot afegir blocs nous a una cadena de blocs, i no es pot modificar el contingut de cap bloc ni eliminar cap bloc després que s'afegeixi a la cadena de blocs.
>
> Cada bloc emmagatzema un *hash* criptogràfic del bloc anterior com a enllaç. Una funció *hash* criptogràfica és una funció que pren dades com a entrada i retorna una cadena de caràcters de longitud fixa que representa les dades d'entrada.
>
> El valor retornat es diu valor de resum. Si realitza algun canvi en les dades d'entrada i torna a executar la funció *hash*, obtindrà un valor de resum diferent. Els valors *hash* faciliten la detecció de canvis en els blocs de *blockchain*. Qualsevol canvi en un bloc trenca la cadena a l'invalidar l'enllaç del següent bloc[7].

---

[5] Traduït com a P2P, o node a node (*peer-to-peer*) o xarxa entre iguals. Element clau per entendre la *blockchain*. És el factor que permet la descentralització. Un node té una doble funció (client-servidor) de consumir o d'abastir informació. Si un d'aquest nodes pogués fallar /eliminar-se, restarien els altres per poder transmetre-la (Hausser, 2006).

[6] Formalitzat mitjançant el *smart contract*, un contracte que estableix les regles del joc entre els usuaris que hi participen, decidint que fa i que no pot fer cada usuari, i que és "un programa d'ordinador que verifica i executa les seves condicions quan es produeixen esdeveniments predeterminats. Un cop codificat i introduït a la cadena de blocs, el contracte no es pot canviar i funciona d'acord amb les seves instruccions programades" (Giancaspro, 2017).

[7] Per exemple si s'utilitza una funció *hash* SHA256 (*secure hash algorithm*, algoritme criptogràfic de seguretat de 256 bits, força utilitzada en criptomoneda) i s'introdueix un text com per exemple "*Universitat Oberta de Catalunya*", la funció retorna un text de mida fixa, en aquest cas 64 lletres que conté la informació de manera encriptada:

829BA3D8032BAA5FA0843038F2FBD1D1DE878604839F916097AC729EE3D0E7CF

Si es modifica encara que sigui un valor, per tant canviant el contingut - i seguint amb l'exemple anterior - traient-li la lletra U a Universitat, *"niversitat Oberta de Catalunya"*, el valor *hash* canvia i no coincideix amb l'anterior *hash:*

489F2DB23DE70B397B812183638CDC263C31A452F7780C36ECD130E6717545B8

Cada bloc de la cadena emmagatzema el valor de resum original del bloc anterior. Els canvis en qualsevol bloc fan que el valor de resum d'aquest bloc canviï, el que significa que el valor de resum emmagatzemat en el següent bloc ja no coincideix amb el valor de resum del bloc actual. Qualsevol node[8] pot saber ràpidament si algun bloc va canviar des que es va agregar.

Quan un nou node complet s'uneix a la xarxa *blockchain*, descarrega una còpia de tots els blocs que es troben actualment a la cadena[9]. Una vegada que el nou node es sincronitza amb els altres nodes i té l'última versió de la cadena de blocs, rep els nous blocs, a l'igual que altres nodes." (Solomon, 2019, pàg. 6-7).

[8] Són - anomenats també servidors informàtics - els que serveixen informació a altres ordinadors (Houben i Snyersc, 2018). Hi trobem els anomenats nodes complets (*full nodes*) que "emmagatzemen una còpia completa de [tota] la cadena" i els nodes lleugers (*lightweight nodes*) [que] només emmagatzemen els blocs més recents i poden sol·licitar blocs més antics sota comanda quan els usuaris els necessiten" (Solomon, 2019).

[9] Per garantir la correcta seqüenciació de les transaccions "qualsevol node d'una xarxa de *blockchain* pot proposar l'addició de nova informació a la cadena, [i per tant] els nodes han d'arribar a alguna forma d'acord" (Houben i Snyersc, 2018). Aquest forma d'acord s'estableix mitjançant els mecanismes o algorismes de consens, que són algorismes matemàtics, mètodes de validació algorítmica predefinits, que els utilitzem quan volem afegir o determinar col·lectivament la validesa del nou bloc. Al treballar amb un sistema P2P - al no tenir una figura autoritzada que prengui les decisions, es necessita arribar al consens. Per resoldre aquest consens s'utilitza el problema de lògica anomenat *Byzantine Generals Problem,* on si trobem una ordre (o un mandat) contradictori es donarà per "verdadera l'ordre majoritària" (Zamorano, 2018).

Un exemple de mecanisme de consens és la PoA, acrònim de prova d'autoritat, que utilitza "la identitat com a garantia per obtenir el dret a validar blocs. La validació dels blocs deixa de ser un procediment anònim, de manera que les identitats i les reputacions públiques dels validadors esdevenen incentius per garantir el comportament honest" (i2CAT, 2020).

2.2. Deepfake

Al 2020, i realitzat pel MIT *Center for Advanced Virtuality*, es presentava el projecte audiovisual *In Event of Moon Disaster* on es recrea un *fake* complet del president Nixon. En aquest fals-document historiogràfic es "reescriu un moment important de la història per educar el públic sobre els perills dels *deepfakes*" (MIT Open Learning, 2020). L'àudio i els llavis (*fake lyp-sinc*) de Nixon havien estat manipulats, anunciant que els astronautes de l'*Apollo 11* no havien pogut tornar a casa. Però no era del tot fals els discurs. Existeix un document, escrit l'any 1969, que recull el "discurs presidencial preparat per si els astronautes del mòdul lunar no podien eixir dels estranys lligams de la lluna i quedar-se allà". (Sorkin, 2017).

Això és exactament el que provoca la visualització del *deepfake*, entre indeterminació i cinisme, i més enllà "d'actituds postmodernistes o nihilistes" la veritable amenaça és tenir "dificultats per creure el que els ulls o orelles ens estan dient, fins i tot quan la informació és real" i a arribar a pensar en que la veritat què se'ns presenti pugui ser falsa (Chesney i Citron, 2019).

Però "les manipulacions als mitjans [audiovisuals] són tan antigues com els mitjans mateixos"[10] (Kietzmann *et al.*, 2020). En la història de la manipulació hi han hagut diverses tècniques ja sigui en un intent de *damnatio memoriae*, o bé amb finalitats artístiques, comercials o bé polítiques. Ja al llarg del segle XX, es fan grans avenços i des dels anys 50 van apareixent eines de *computer graphics*, fins als anys 80 que es comencen a comercialitzar les *raster graphics editors* com la popular eina *Adobe Photoshop*.

Però per valorar el gran impacte social molts autors considerant que l'AI i en concret la tecnologia associada a les *deepfakes* "ha marcat un punt d'inflexió en el creació de contingut fals" i que actualment "s'ofereixen procediments automatitzats per crear contingut fals que són impossibles "de detectar per als observadors humans" (Westerlund, 2019).

[10] De fet es considera la primera manipulació fotogràfica, d'un calotip o talbotip, l'any 1846, quan en una foto de grup es va fer desaparèixer una persona ja que trencava "la composició de l'escena i es va pintar, al negatiu, la figura [del personatge] amb tinta". Article d'Ann Young. *History of Photo Editing* (1826-2019). Consultat a https://bit.ly/3gg2Z4C

Com que les *deepfakes*, actualment, estan basades en ML, aquestes entrenen[11] i amb quantitats molts *datasets*, i per aquest motiu el fenomen es va donar conèixer amb actors i *celebrities*, per la quantitat de dades en línia i d'alta qualitat que es podia trobar d'aquestes persones. A hores d'ara els *datasets* els "ofereix" qualsevol usuari que comparteix la seva identitat en línia, amb o sense el seu consentiment.

No hi ha exactament una data exacta de l'aparició del fenomen on poder trobar consens, però podria situar-se entre la publicació dels següents articles i esdeveniments:

- La publicació de l'article "Generative Adversarial Networks" de Goodfellow *et al.*, (2014), que presentava les xarxes generatives antagòniques (en anglès, *generative adversarial network*), mitjançant dos models generatius (dues xarxes neuronals) en un procés contradictori, entrenen simultàniament: el primer model el generatiu "que captura la distribució de dades" i el segon model o discriminatiu "que estima la probabilitat que una mostra provingui de les dades" de creació en lloc de les del model generatiu. És busca que "el model discriminador cometi un error" i per tant es creïn noves (o opcions) de dades que no corresponen (o que son derivades de les d'entrada)[12].

[11] "el procés amb el qual la IA aprèn per crear un model". Article de Josep Roca, *¿Qué significan Entrenamiento e Inferencia en la Inteligencia* Artificial?. Consultat a https://bit.ly/351fjAg

[12] Per esclarir una altre manera de definir-ho podria la que exposa wiki.org on aquestes dues xarxes neuronals artificials es van entrenant competint entre elles:

> "La primera xarxa, anomenada generadora, té la tasca de crear dades falsificades, com ara fotos, enregistraments d'àudio o imatges de vídeo, que repliquen les propietats del conjunt de dades original. La segona xarxa , la discriminadora, té la tasca d'identificar les dades falsificades. En funció dels resultats de cada iteració, la xarxa de generadors s'ajusta per crear dades cada vegada més realistes. Les xarxes continuen competint (sovint per milers o milions d'iteracions) fins que la generadora millora el seu rendiment de manera que la discriminadora ja no pugui distingir entre dades reals i falsificades". (Wikia.org)
>
> Consultat a https://itlaw.wikia.org/wiki/Deep_fake

- La publicació de l'article "Face2face: Real-time Face Capture and Reenactment of RGB vídeos". (Thies *et al.*, 2016)[13], on s'animaven les expressions facials d'un vídeo de destinació (persona nº1) mitjançant un actor (persona nº2). Després es tornava a renderitzar i el resultat era la confecció d'un vídeo fotorealista manipulat (la persona nº1 s'expressava i movia els llavis com la persona nº2).

- La publicació de l'article "Synthesizing Obama: Learning Lip Sync from Audio" (Suwajanakorn *et al.*, 2017), on mitjançant l'arquitectura LSTM (*long short-term memory*) que empra les RNN (xarxes neuronals recurrents, o *recurrent neural network* en anglès), amb només un àudio com a font (com a input d'entrada, a diferència de Thies *et al.*, 2016) van sintetitzar un vídeo d'alta qualitat amb Barack Obama, parlant amb sincronització de llavis precisa. Això també es va aconseguir entrenant durant moltes hores de les imatges d'Obama, i on "una xarxa neuronal recurrent aprenia el mapatge" associant les pistes d'àudio "fins a formes de boca".

- A la comunitat *Reddit* un usuari anònim es fa passar amb el sobrenom de '*Deepfake*' (*deep* d'aprenentatge profund, i *fake* de falsificacions) i comparteix els primers vídeos hipertrucats col·locant actrius en videoclips per a adults.

Com assenyala Masood *et al.* (2021), des del 2014 fins a l'actualitat, hi ha hagut un creixement en els avenços i en la millora molt substancial de "la qualitat de les cares sintètiques, generades per les variacions en les GAN". Aquest mateixos autors enfoquen les eines de detecció, en relació a cinc tipologies des de la *Face-Swap* (intercanvi de cares), el *Lip-Sync* (sincronització de llavis), *Face-Reenactment* (també conegudes com *Puppet-Master deepfake*, *deepfakes*

[13] Malgrat sembla una novetat, Bregler *et al.* (1997), ja presenten un model de resultats similars, però de menys qualitat, gairebé dues dècades abans.

titellaires), imatges generades per GAN[14] i *fakes* d'àudio. També es podria optar per una classificació similar com aquesta: "fotografia *(face- and body-swapping),* àudio *(voice-swapping, text to speech),* vídeo *(face-swapping, face-morphing, full body puppetry)* i vídeo amb àudio-vídeo *(lip-synching)*" (Kietzmann *et al.*, 2020).

La *deepfake* encaixa per tant amb aquesta definició d'àudio i també de vídeo que segons Albahar i Almalki (2019) citant a Korshunov i Marcel (2018), és defineix com "una tècnica de manipulació que permet a un usuari canviar la cara d'un individu, sovint un actor, actriu o qualsevol altra celebritat amb qualsevol altre actor o persona. [Son per tant] imatges, vídeos i àudio falsos que semblen i/o sonen autèntics".

No s'han d'obviar els avantatges de les falsificacions amb *deep learning* i l'ús de l'AI, perquè poden ser positives incloent diversos camps com el de l'experimentació artística[15] (Chesney i Citron, 2019).

Però genera preocupació i alarma, per les utilitzacions i en combinació de futurs escenaris de tipologia *Cambridge Analytica*, de desprestigi personal, de ciberxantatge o bé per el consum cada cop més al alça de notícies online o bé per xarxes socials. Per aquests motius hi hauria implicacions de tipus social, polític i econòmic i altres situacions que se'n puguin derivar[16].

Malgrat això, tant a Farid (2021) com per Masood *et al.* (2021) no hauríem de confondre - tant a nivell periodístic com a la literatura acadèmica - les *deepfakes* amb les "low-tech videos", "cheapfakes" o "shallow fakes", o altres de *retouching*

[14] En la literatura consultada com a nivell divulgatiu del fenomen es conegut a la pàgina web "this person doesnt exist" basat en el paper de Karras, et al., (2019). Analyzing and improving the Image Quality of Stylegan.

[15] podrien oferir l'exploració de *what-if situations*, situacions que mai van existir o obres que mai van existir Aquí s'inclourien projectes per escoltar nova música de músics desaperaguts (projecte benèfic anomenat *Lost Tapes of the 27 Club*); descobrir noves pintures de pintors d'altres èpoques (projecte de *Microsoft, The Next Rembrandt*); recreació de la veu i cant (*WaveNet* o el *SV2TTS*, clonador de veu a partir d'una petita mostra); projectes de *digital necromancy*; GPT-3 o Grover (Generador d'AI de text d'un abast i potència inimaginable); 'coloritzadors', restauradors, cercadors i generadors d'imatges que mai han existit (DALL-E), interpretadors 2D -3D (NeRF) aquest darrers ja obririen un debat per si sols en el terreny de la documentació audiovisual.

[16] Com suggereix Engler (2020), els efectes seran triples: [1] desinformació (i sensació d'Efecte Mandela, coses que no han succeït però les recordem com a tal), [2] esgotament del pensament crític (provocació d'incertesa) i [3] El dividend del mentider (*the liar's dividend*, concepte exposat a Chesney i Citron, 2019): "la forma en què un entorn on no està clar què és real i què és fals pot beneficiar a aquells que creen i difonen falsificacions[...] a més d'alimentar les flames de les falsedats, els esforços de desmentiment legitimen el debat sobre la veracitat". (Macmillian Dictionary)

(*photoshopped*) que serien "vídeos generats amb baixa tecnologia" o "manipulacions audiovisuals creades utilitzant eines d'edició bàsica que són més barates i de programari més accessible [...] mitjançant l'alentiment, l'acceleració, el tall i la combinació selectiva d'imatges inalterades existents".

Fins ara, aquests *cheapfakes* eren relativament fàcil fer-ne una detecció ja sigui per la sensació d' *uncanny valley*[17] (Cherry i Lustik, 2020), o per mètodes com l'anàlisi de la comprensió de la imatge i la relació de píxels aplicant la llei de Benford *o Newcomb–Benford law* (Brigden, amb Hany Farid, 2020), o per l'ús d'eines com la CBIR (*content-based image retrieval*), les CBVIR o les QBIC eines de cerca conegudes popularment com *reverse image* que ofereixen els navegadors com *Google* o *Yandex*, eina que encara podria tenir un llarg recorregut, i que ofereixen la recuperació d'imatges o informació basada en el contingut (Marques i Furht, 2002).

A Tewari *et al.* (2020), també insisteixen en les implicacions socials d'aquestes innovacions (així com també de la velocitat que ha après la tecnologia associada com les *deepfakes*), i de possibles elements de detecció com a contra mesura com són la detecció passiva automàtica d'imatges sintètiques o manipulades; ofereixen una de divisió en 2 gran corpus de literatura forense sobre mitjans digitals dividits en manipulació-específica i manipulació-independent. També proposen mètodes de protecció proactiva (com signatures digitals, marques d'aigua o índex biomètrics) i una anàlisi forense passiva.

La preocupació ha tingut tant impacte, que DARPA, l'Agència de Projectes de Recerca Avançada de la Defensa, del Departament de Defensa dels Estats Units ha engegat dos projectes com *SemaFor* (*Semantic Forensics*) i *MediaFor* (*Media Forensics*) amb el que vol detectar les manipulacions en aquest els camps de la informació falsa i la imatge generada per AI. Les línies d'investigació ara estan encarades en les inconsistències de tipus lumínic (Corrigan, 2021). També les xarxes i plataformes socials estan posant fil a

[17] Traduït com a vall inquietant, sensació d'estranyesa o poca naturalitat en observar els moviments de la robòtica o del món de l'animació. El concepte fou introduït per Masahiro Mori l'any 1970.

l'agulla, amb la competició anomenada DFDC (*DeepFake Detection Challenge*) de *Facebook*, on s'hi poden presentar diversos projectes de detecció i que ha "permès que experts de tot el món es reuneixin, comparin els seus models de detecció de *deepfake*, i provin nous enfocaments" (Canton Ferrer *et al.*, 2020). Per fer-ho les *tech companies* han compartit*, datasets*[18] per fer les els assajos i entrenar els models.

Ara el problema no radica tant en els innovacions relacionades amb les GAN, o l'AI en general, sinó més aviat amb els algorismes de recomanació, ja que promouen de manera agressiva que les *deepfake* o les *fake news* apareguin "les fonts de notícies i a les llistes de visualització, que submergeixen els usuaris en cambres de ressò cada vegada més aïllades i sense realitat" (Farid, 2021).

Identificant doncs la problemàtica, Farid[19] (2019), intuïa el que podria ser el millor mètode per autenticar: "quan graveu algun esdeveniment notable, i si no voleu que la gent posi en dubte l'autenticitat del vostre vídeo o de la vostra imatge, [...] i en lloc de capturar amb una càmera *iPhone* o *Android* estàndard, utilitzeu un programari de captura controlada [*controlled capture software*]. [...] En el moment de gravar, signeu el contingut criptogràficament [*cryptographically sign content*]. Ho poses a la cadena de blocs, [què és] un llibre major distribuït i immutable. Aleshores podreu, amb una confiança bastant alta, autenticar el contingut que passa per la línia. Potser és per aquest camí on hem d'anar."

[18] un conjunt de dades, en aquest cas un "gran conjunt de dades que contenen enregistraments, vídeos o fotos" manipulades i d'altres pristínes (Albahar i Almalki, 2019).

[19] enginyer informàtic i professor de la Dartmouth College, assessor DARPA, guru i conegut com el "pare" (anomenat així per *Nova*) de la *digital image forensics o digital forensic,* divulgador i expert amb un llarg recorregut aquest camp. Consultat a https://www.pbs.org/wgbh/nova/sciencenow/0301/03.html

3. METODOLOGIA

3.1. Tipologia i estratègia d'estudi

Per a la realització d'aquest treball final de grau s'ha realitzat una revisió bibliogràfica, acotada al període de temps que comprèn des de gener de 2017 fins al maig de 2021, de les principals fonts i bases de dades a què té accés la pròpia UOC i d'altres d'externes, algunes d'elles són especialitzades en informàtica i en ciències de la informació: *Association for Computing Machinery,* DIALNET*,* DOAJ*, Google Scholar Acadèmic*, IEEE *Xplore, Microsoft Academic, ProQuest Central, ScienceDirect, Springer Link,* TDX, i *Web of Science*. També ha ajudat tenir l'extensió al navegador de la *Lean Library*, de SAGE, ofert per la Biblioteca de la UOC, que ha permès detectar ràpidament articles i bibliografia complementària que es troba a disposició de la UOC, reben notificacions emergents en el moment de consultar en els cercadors més habituals.

S'ha triat els articles i literatura més rellevants publicats en els darrers anys. Es podria haver triat a partir de les dates de les publicacions de Goodfellow *et al.* (2014), o la data de Thies *et al.* (2016). La data escollida, però, és el 2017, l'any en què es dona a conèixer el fenomen (La polèmica publicació a Reddit i el projecte "The Synthesizing Obama"), sota el concepte de *deepfake*.

Dins d'aquest interval, s'ha escollit la literatura acadèmica que posa més èmfasi en el *blockchain* com a solució per a la protecció de material gràfic i audiovisual, i no tant en la vessant de detecció (*detection*). L'idioma que s'ha triat ha estat fonamentalment l'anglès, ja que és l'idioma predominant en l'actual marc científic de la intel·ligència artificial (AI) i els seus subcamps com el de la *machine learning* (ML) i la *deep learning* (DL).

També caldria incidir en el fet que pot semblar que el conjunt d'articles finals sigui reduït, però ha estat una selecció acurada, precisa i actualitzada, seguint els criteris en aquest camp com assenyala Centobelli *et al.* (2021), en el seu estudi bibliomètric. Per exemplificar-ho, es podria esmentar el cas de Bui *et al.* (2019) i el seu model de DLT anomenat *Archangel* on empren *hashs* de contingut temporal a la cadena de blocs, amb similituds respecte el treball de

Hasan i Salah (2019); en aquest cas concret ens hem decidit per aquest darrer per estar més citat.

S'han combinat les paraules clau amb els operadors i al ser un neologisme es fàcil trobar el concepte de *deepfake* amb les següents variants: *deep fake*, *deepfakes*, *deep fakes o deepfake*. La majoria de repositoris no han estat sensibles, ja que han identificat totes les variants com a "deepfake". S'ha utilitzat els operadors booleans i aquests han estat AND, OR, NOT.

Per tant, les paraules clau utilitzades per aquesta recerca han estat: *blockchain* i *deepfake*.

3.1.1. Fonts d'informació consultades

- **Association for Computing Machinery**

  Mitjançant subscripció de la UOC, es consulta el repositori de l'associació d'informàtica, que està considerada com la més gran del món, amb els termes de cerca *deep fakes* AND *blockchain*. S'obtenen 833 resultats i triem els que ordena ACM per rellevància.

- **DIALNET**

  Aquest portal bibliogràfic, referent de la producció de literatura científica hispana, no ha donat cap resultat.

- **DOAJ**

  Es consulta aquest conegut repositori en obert multidisciplinari d'articles i revistes científiques. S'obtenen 2 resultats, però resulten coincident amb altres cerques.

- **Google Scholar Acadèmic**

  S'utilitza el repositori digital propi de *Google*, el *Google Scholar Acadèmic*. Amb els termes de cerca *deep fake* AND *blockchain*, obtenim 282. Amb els termes de cerca *deepfake* AND *blockchain*, s'obtenen 388.

Amb els termes de cerca *deep fakes* AND *blockchain*, s'obtenen 304 que resulten se coincidents amb les altres dos opcions de cerca. Filtrant per articles i revistes des del 2017 al 2021. S'obtenen 2 resultats i triem els que estiguin en obert i els que ordena GSA per rellevància.

- **IEEE Xplore**

  Mitjançant subscripció de la UOC, es consulta amb els termes de cerca *deepfake* OR *deep fake* AND *blockchain* NOT *detection*, i s'utilitza la biblioteca digital i base de dades d'articles, actes, conferències, i altra literatura grisa de temàtica específica d'informàtica i enginyeria del IEEE Xplore, on s'hi poden trobar diferents tipus de documentació com llibres, conferències, cursos i revistes. Els termes de cerca són *deepfake* AND *blockchain*. Filtrant per articles i revistes, S'obtenen 132 resultats.

- **ProQuest Central**

  Mitjançant subscripció de la UOC, es consulta aquesta base de dades de recerca multidisciplinària, amb els termes de cerca *deepfake* AND *blockchain*. S'obtenen 17 resultats i triem per tipologia d'articles revisat per parells i els que ordena PQC per rellevància.

- **ScienceDirect**

  Mitjançant subscripció de la UOC, es consulta la col·lecció d'Elsevier de publicacions de ciències físiques i enginyeria, amb els termes de cerca *deep fake* AND *blockchain*, s'obtenen 8 resultats i amb els termes de cerca *deepfake* AND *blockchain*, s'obtenen 17 resultats. Triem els que ordena SD per rellevància.

- **Springer Link**

  Mitjançant subscripció de la UOC, s'accedeix a diversa documentació científica de revistes, llibres, i altres tipologies. Amb els termes de cerca *deepfake* AND *blockchain*. S'obtenen 20 resultats i triem els que ordena SL per rellevància.

- **Microsoft Academic**

  S'utilitza el cercador de literatura científica de Microsoft. S'obtenen 266 resultats i triem els que estiguin en obert i els que ordena MA per rellevància.

- **TDX**

  El repositori que conté les tesis doctorals en format digital gestionat pel CSUC, no ha donat cap resultat.

- **Web of Science**

  Mitjançant subscripció de la UOC, es consulta aquesta base de dades de contingut científic, gestionada pel FECYT. S'obtenen 2 resultats, però resulten coincident amb altres cerques.

3.1.2. *Flowchart* o taula/resum

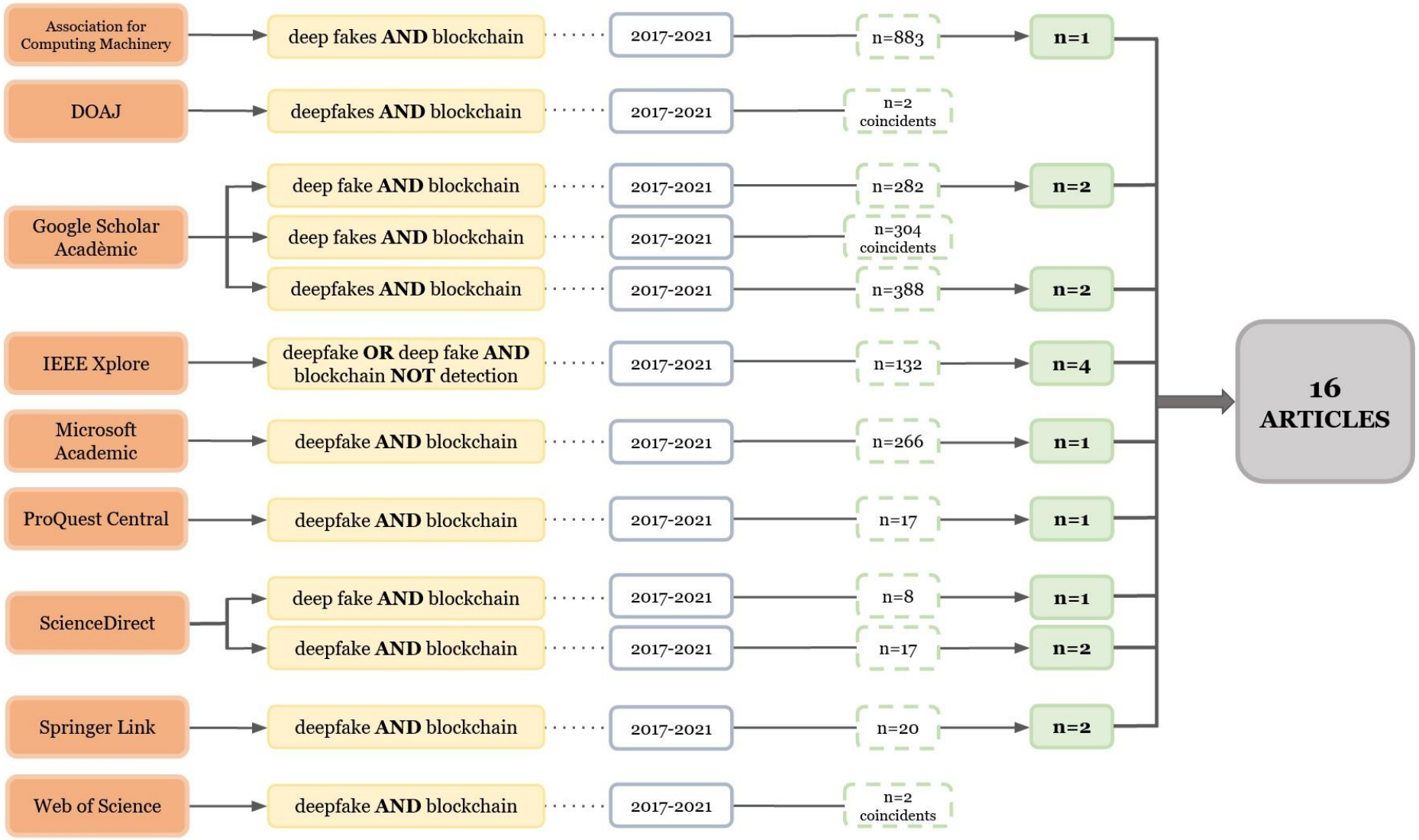

Figura 1. Flowchart dels articles seleccionats. Font: Elaboració pròpia.

4. RESULTATS

A continuació es presenten la taula que presenta els 16 articles seleccionats ordenats per any de publicació i per ordre alfabètic, amb un resum del format IMRaD (introducció, metodologia, resultats i discussió).

Taula 1. Resultats dels articles seleccionats. Font: Elaboració pròpia.

AUTORS	ANY	TÍTOL	TIPOLOGIA	DISCIPLINA	INDEXACIÓ CONTROLADA/ KEYWORDS	REVISTA/ CONFERÈNCIA/ PUBLICACIÓ	RESUM IMRaD
Hasan i Salah	2019	"Combating Deepfake Videos Using Blockchain and Smart Contracts"	Article de recerca	Computer Science	contracts cryptography meta data video signal processing learning (artificial intelligence)	IEEE Access	Es proposa una solució de plataforma distribuïda *Etherum* i la proposta de cadena de blocs com aprova d'autenticitat de vídeo digitals; es visualitza l'arquitectura del sistema, i altres detalls de disseny d'entitat-relació, diagrames i algoritmes que formulen les interaccions i transaccions. Mostra la seva implementació, amb ús d'emmagatzematge descentralitzat, analitza la despesa associada i el grau de seguretat assolida. Aquesta solució és aplicable altres documents digital, rastrejant la font i identificant-la com a confiable i acreditada.

Taula 2. Resultats dels articles seleccionats. Font: Elaboració pròpia.

AUTORS	ANY	TÍTOL	TIPOLOGIA	DISCIPLINA	INDEXACIÓ CONTROLADA/ KEYWORDS	REVISTA/ CONFERÈNCIA/ PUBLICACIÓ	RESUM IMRaD
Qayyum *et al.*	2019	"Using Blockchain to Rein in the New Post-Truth World and Check the Spread of Fake News"	Article de recerca	Computer Science	cryptography information dissemination information filtering learning (artificial intelligence) social networking (online)	IT Professional	Es fa un anàlisi descriptiva d'un entorn de treball basat en la cadena de blocs amb contractes intel·ligents per fer front a notícies falses de qualsevol naturalesa. Enumera les característiques i avantatges de la *blockchain*. Un cop presentat el marc teòric de reducció i detecció de *fake news*, d'aquesta solució pròpia, obre la porta a poder-ne dissenyar un prototip.
Shae i Tsai	2019	"AI Blockchain Platform for Trusting News"	Article de recerca	Computer Science	cryptocurrencies distributed databases graph theory Internet social networking	International Conference on Distributed Computing Systems	És descriu una arquitectura de sistema d'una cadena de blocs amb intel·ligència artificial per a construir una forta col·laboració entre els investigadors de la *blockchain* i els mitjans de comunicació per avançar en la investigació que lluita contra les notícies falses. És raonen les variables que participen en un ecosistema de notícies contrastades amb *blockchain*, així com els mecanismes de classificació i modelització.

Taula 3. Resultats dels articles seleccionats. Font: Elaboració pròpia.

AUTORS	ANY	TÍTOL	TIPOLOGIA	DISCIPLINA	INDEXACIÓ CONTROLADA/ KEYWORDS	REVISTA/ CONFERÈNCIA/ PUBLICACIÓ	RESUM IMRaD
Westerlund	2019	"The Emergence of Deepfake Technology: A Review"	Article teòric	Computer Science	artificial intelligence cybersecurity deep learning Deepfake fake news	Technology Innovation Management Review	Es fa un anàlisi de n=84 articles sobre notícies que tracten el tema de les *deepfakes* per examinar la tendència del fenomen, els seus pros i contres, i els processos de com és combaten. Es proporciona una revisió completa per detectar les oportunitats futures comercials, en la lluita en contra de les *deepfakes* i les *fakenews*. Es destaca la possible valoració d'aportacions i opinions personals en fòrums, d'artistes digitals, desenvolupadors informàtics o la comunitat en general.
Alattar *et al.*	2020	"A System for Mitigating the Problem of Deepfake News Videos Using Watermarking "	Article de recerca	Computer Science	Authentication Blockchain Deep Learning DeepFakes Fake news Watermarking	Electronic Imaging	Utilitzant les marques d'aigua digital de l'empresa *Digimarc*, l'estudi proposa integrar-les a les pistes d'àudio i vídeo, o bé en el moment en què es capturen els vídeos o després durant la seva distribució. Aquesta marca d'aigua, ara vinculada a la còpia original i a la seves metadades, a dins de la cadena de blocs, se'n podria fer-ne l'anàlisi forense del vídeo o imatge digital.

Taula 4. Resultats dels articles seleccionats. Font: Elaboració pròpia.

AUTORS	ANY	TÍTOL	TIPOLOGIA	DISCIPLINA	INDEXACIÓ CONTROLADA/ KEYWORDS	REVISTA/ CONFERÈNCIA/ PUBLICACIÓ	RESUM IMRaD
Fraga-Lamas i Fernández-Caramés	2020	"Fake News, Disinformation, and Deepfakes: Leveraging Distributed Ledger Technologies and Blockchain to Combat Digital Deception and Counterfeit Reality"	Article de revisió	Computer Science	data privacy Internet peer -to- peer computing security of data social networking	IT Professional	Fa una descripció general de com fer front a les noticies falses i *deepfakes* mitjançant les DLT o tecnologies de registre distribuït, i també enumera les diferents aplicacions DLTs que podem trobar actualment per combatre-les. Exposa les limitacions logístiques i tecnològiques de les DLTs, però les destaca per ser les que millor garantitzen l'autenticitat, la comprovació i eliminació de contingut falsejat.
Jaroucheh *et al.*	2020	"TRUSTD: Combat Fake Content Using Blockchain and Collective Signature Technologies"	Article de recerca	Computer Science	cryptography data privacy digital signatures Internet learning (artificial intelligence) social networking Web sites	Computer Software and Applications Conference	Es presenta una descripció general de TRUSTD, un ecosistema basat en la cadena de blocs, amb signatura de grup o col·lectiva, per combatre el contingut fals. Aquesta eina permetrà jutjar la credibilitat d'aquest mateixos continguts, entre els usuaris participants seran creadors de continguts, i altres actors de valoració, com poden ser notícies sorgides del periodisme ciutadà.

Taula 5. Resultats dels articles seleccionats. Font: Elaboració pròpia.

AUTORS	ANY	TÍTOL	TIPOLOGIA	DISCIPLINA	INDEXACIÓ CONTROLADA/ KEYWORDS	REVISTA/ CONFERÈNCIA/ PUBLICACIÓ	RESUM IMRaD
Ki Chan *et al.*	2020	"Combating Deepfakes: Multi-LSTM and Blockchain as Proof of Authenticity for Digital Media"	Article metodològic	Computer Science	blockchains feature extraction Internet learning (artificial intelligence) meta data recurrent neural nets social networking	International Conference on Artificial Intelligence	Es descriu un marc teòric de cadena de blocs descentralitzada de vídeo i fotografia digitals, que utilitza les LSTM com codificador per crear funcions discriminatòries úniques, mitjançant un *hash*, una funció criptogràfica, que emmagatzema els subtítols descriptius i característiques del contingut de les imatge/s.
Kietzmann *et al.*	2020	"Deepfakes: Trick or treat?"	Article teòric	Computer Science	Deepfakes Fake news Artificial intelligence Machine learning Deep neural networks	Business Horizons	Descriu un marc per a la gestió i disminució dels riscos vers els *deepfakes*, anomenat REAL. Aquest marc està basat en quatre punts que haurien de ser el de preservar el contingut original per garantir la negació; denunciar i exposar les *deepfakes* tant bon punt es detectin; creació d'un marc legal específic que afavorís la protecció legal, i com a darrer punt que els grans marques i companyies tinguessin un compromís ètica fort per afavorir la confiança. El contingut preservat pot utilitzar empremtes digitals, al mateix temps que es produeix, que es grava, com *Amber Authenticate*.

Taula 6. Resultats dels articles seleccionats. Font: Elaboració pròpia.

AUTORS	ANY	TÍTOL	TIPOLOGIA	DISCIPLINA	INDEXACIÓ CONTROLADA/ KEYWORDS	REVISTA/ CONFERÈNCIA/ PUBLICACIÓ	RESUM IMRaD
Mitra *et al.*	2020	"A Machine Learning Based Approach for Deepfake Detection in Social Media Through Key Video Frame Extraction"	Article de revisió	Computer Science	Deepfake Deep learning Key video frame extraction Depthwise separable convolution neural network (CNN) Transfer learning Social media Compressed video	SN Computer Science	Es descriu una metodologia basat en l'ús de xarxes neuronals per a la detecció de vídeos falsejats. Aquest model s'integra d'element com una xarxa neuronal convolusional (CNN), una xarxa de classificació i un algoritme, que redueix la càrrega i ús computacional. Empra la xarxa *Xception*, en combinació amb aplicacions de detecció i datasets com *Deepfake Detection* o *FaceForensics* ++. S'afirma un nivell de detecció per sobre del 90%, mitjançant la extracció de fotogrames clau, malgrat que cal reduir la memòria d'execució i la mida del model.
Skibba	2020	"Accuracy Eludes Competitors in Facebook Deepfake Detection Challenge"	Revisió narrativa	Computer Science	Science Journalism	Engineering	La revisió copsa les darreres notícies en l'aposta de les grans empreses com *Facebook* en la detecció de *deepfakes*. I exposa el projecte DARPA que contempla la utilització *blockchain*.

Taula 7. Resultats dels articles seleccionats. Font: Elaboració pròpia.

AUTORS	ANY	TÍTOL	TIPOLOGIA	DISCIPLINA	INDEXACIÓ CONTROLADA/ KEYWORDS	REVISTA/ CONFERÈNCIA/ PUBLICACIÓ	RESUM IMRaD
Yazdinejad *et al.*	2020	"Making sense of blockchain for AI deepfakes technology"	Article teòric	Computer Science	blockchains deep learning (artificial intelligence)	Global Communications Conference	L'objectiu d'aquest paper es poder aplicar el BK contra els vídeos DF. Estudi que posa el focus en el procés d'autentificació amb l'AI, i on pot ser útil com actor principal la cadena de blocs. Aporta casos d'ús, defineix els pros i contres de cada d'un d'ells i exposa reptes futurs per a la investigació en la relació *blockchain-deepfakes*.
Agrawal *et al.*	2021	"DeHiDe: Deep Learning-based Hybrid Model to Detect Fake News using Blockchain"	Article de recerca	Computer Science	Blockchain Provenance Deep Learning Fake News Detection Proceedings	International Conference of Distributed Computing and Networking	Es proposa DeHiDe, un model híbrid de *Deep Learning*, que combinarà la tecnologia *blockchain* per combatre les *fake news*. L'arquitectura del sistema, amb smart contracts, estarà formada per tres perfils d'usuari: el Reporter, l'Analitzador (de tipus DL i periodistes reals) i el Validador. S'atorgaran qualificacions (punts de credibilitat). Obra la porta perquè es pugui analitzar l'impacte de la seva implementació

Taula 8. Resultats dels articles seleccionats. Font: Elaboració pròpia.

AUTORS	ANY	TÍTOL	TIPOLOGIA	DISCIPLINA	INDEXACIÓ CONTROLADA/ KEYWORDS	REVISTA/ CONFERÈNCIA/ PUBLICACIÓ	RESUM IMRaD
Centobelli *et al.*	2021	"Surfing blockchain wave, or drowning? Shaping the future of distributed ledgers and decentralized technologies"	Revisió bibliomètrica	Data Science	Bibliometric analysis Block-chain Decentralized technology Distributed ledger Literature review Network analysis Performance analysis Traceability Tracking Transparency trust	Technological Forecasting and Social Change	Proporciona informació sobre la temàtica intentant copsar el creixent interès, amb el llarg potencial que té la tecnologia *blockchain* en diversos camps. Amb l'estudi de n=2233 publicacions. Exposa el paper fonamental a la privacitat limitada en el temps a *blockchain* i la privacitat transaccional. Obrint el pas cap a l'evolució de la tecnologia *blockchain* 4.0
Jing i Murugesan	2021	"Protecting Data Privacy and Prevent Fake News and Deepfakes in Social Media via Blockchain Technology"	Article de revisió	Computer Science	Data privacy Fake news Deepfake Blockchain	Communications in Computer and Information Science	Es proporciona una revisió del impacte de les *fake news* i els *deepfakes*, sobretot en les xarxes socials. Advoca per la tecnologia de cadena de blocs i proposa d'un nou model d'índex de confiança per assolir la integritat de les dades i la privacitat dels usuaris. El model d'índex de confiança treballarà com a mecanisme per rastrejar la font primigènia de notícies falses i *deepfakes*, avaluant l'autenticitat i confiabilitat de l'autoria. El model teòric obre el camí per crear-ne un prototip per desenvolupar-lo i comercialitzar-lo.

Taula 9. Resultats dels articles seleccionats. Font: Elaboració pròpia.

AUTORS	ANY	TÍTOL	TIPOLOGIA	DISCIPLINA	INDEXACIÓ CONTROLADA/ KEYWORDS	REVISTA/ CONFERÈNCIA/ PUBLICACIÓ	RESUM IMRaD
Masood *et al.*	2021	"Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward"	Article de revisió	Computer Science	Cryptography Security Machine Learning Sound Audio Speech Processing Image Video Processing	arXiv: Cryptography and Security	S'exposa una revisió i anàlisi exhaustiva d'eines existents i enfocaments basats en aprenentatge automàtic (machine learning) de generació *deepfakes*, així com diverses metodologies emprades per detectar-les, tant en àudio com en vídeo. Enumera llums i ombres d'aquestes eines de detecció, apostant per enfocaments híbrids de detecció i protecció, remarcant la manca de literatura científica en detecció d'àudio falsejat, assenyala reptes i nous enfocaments, en aquesta lluita.

5. DISCUSSIÓ DELS RESULTATS

Després d'analitzar la lectura dels diferents articles, on es mostra en tots ells la preocupació del fenomen *deepfake*, passarem a comentar les diferents posicions dels investigadors implicats en el tema i s'analitzaran les relacions entre autors, seguint un ordre de menor a major complexitat en el contingut dels mateixos.

- Per **Yazdinejad *et al.* (2020)**, és el seu objectiu és autenticar, mitjançant l'aplicació del *blockchain* per combatre els vídeos *deepfakes*. Planteja el contracte intel·ligent como una part important *blockchain*, ja que defineix les regles dels blocs i els seus nodes. Si es trenquen les regles dels nodes no es pot operar.

D'aquesta manera Yazdinejad basant-se en Floridi (2018) aporta una "solució que pot incloure un sistema en el qual els usuaris puguin tenir accés a una procedència de dades fiables del contingut digital i poder rastrejar un element en l'historial per demostrar la seva originalitat i autenticitat".

A més a més proposa - com el projecte *Cortex Labs* - utilitzar la cadena de blocs amb programes d'intel·ligència artificial construint-la i utilitzant models d'aprenentatge profund.

Aquest tipus de sistema permet verificar el contingut audiovisual "mitjançant un algoritme segur d'intel·ligència artificial anti-falsificació". Per tal què les persones puguin constatar com es verifiquen els vídeos, algoritme d'AI *antifake* "s'ha d'executar en la cadena de blocs". O bé de manera preventiva - com fa l'empresa *Truepic* amb la col·laboració de Hany Farid - a través de la creació d'una aplicació, que afegeix una marca d'aigua a les imatges, creant un vincle entre aquesta imatge i la seva còpia. Alattar *et al.* (2020), aposten també per *Truepic* i altres marques comercials, com aplicacions independents al procés, com serien *Serelay* i *Prover*.

Els participants que comparteixin el *media material* a la cadena de blocs, no podran canviar o modificar-los, ja que cal el consens de tots per acceptar les possibles modificacions.

Es constata que l'aplicació de les cadenes de blocs, en aquest terreny està a les beceroles, i per tant que les futures directrius de recerca són:

- Integració amb plataforma de xarxes socials de la *blockchain*.

- Extensió per a navegadors (tipus *add-ons* o extensions, o complements afegitó al navegador web).

- Que *The Big Five* (*Microsoft, Apple, Facebook, Amazon* i *Google*) apliquin algorismes anti *deepfake* amb suport de la *blockchain* i que aquests mètodes anti fakes estiguin, integrats en els ginys tecnològics de captació (en *gadgets*, càmeres o *smartphones*) i apliquin com marques d'aigua durant en moment de la captura.

En aquest darrer punt, els autors ja indiquen que les *Big 5 Tech* estan col·laborant amb les universitats i lliura el material que han recopilat de vídeo hipertrucats, a centres de recerca per entrenar-los als models d'AI i trobar nous mètodes de detecció.

L'article defensa la combinació del *blockchain* i la IA però no desenvolupa o dóna pautes de la part d'actuació d'IA durant el procés d'aplicació.

El *core* de l'article però és la utilització de claus públiques i privades per garantir la seguretat de les imatges, exemplificant-ho amb el paradigma d'Alice-Bob de Diffie i Hellman (1976). Per els autors utilitzar la *blockchain* és una bona opció ja que és una eina òptima en l'actualitat per autenticar l'intercanvi de material audiovisual, ja que es beneficia de la funció *hash*, el *smart contract* i la *Proof of Authenticity* (PoA).

- **Fraga-Lamas i Fernández-Caramés (2020)** també es recolza a l'aplicació de les DLT, en general, destacant les seves virtuts recolzant-se en les tesis de Qayyum *et al.* (2019). En la seva pròpia revisió bibliogràfica conclouen i aposten per la integració operativa entre l'AI i la *blockchain*, i destaca "que els esforços actuals de la comunitat investigadora" se estan centrant "en un tipus de falsificació notícies [és a dir, contingut fals verificable], mentre altres males

pràctiques tot just s'estudien", tal i com han indicat altres autors estudiats (Yazdinejad *et al.*, 2020, etc.)

Consideren que la majoria de les propostes que han investigat de detecció d'enganys digitals són basats en *hashs* criptogràfics, sensibles al soroll quan hi ha un canvi de caràcter, un píxel o altres modificacions en el contingut, que com s'ha demostrat, resulta en un *hash* diferent.

Per aquests mateixos autors, les futures plataformes hauran de garantir la seguretat i transparència, cercant equanimitat entre contingut (p. ex., llibertat d'expressió, dret a rebre informació) i protecció de dades personals.

- **Skibba (2020)**, com ha indicat Yazdinejad *et al.* (2020), hi ha una aposta de les *Big Five*, amb el *Deepfake Detection Challenge* (DFDC) on *Facebook* ha compartit els resultats amb la comunitat de recerca, "per impulsar la creació de noves tecnologies innovadores per detectar falsificacions profundes i suports manipulats".

L'autor destaca que els forenses digitals consultats, troben més òptim treballar en establir la prova (autenticar) del que no es fals, en lloc de perseguir-ho, ja que es vol esquivar l'anomenada fal·làcia de la taxa base (o impacte dels falsos positius), en models de detecció que estan donant bona fiabilitat de detecció.

Skibba torna a coincidir amb el plantejament de Yazdinejad *et al.* (2020), esmentant el treball Shweta Jain (2018), quan desenvolupa un projecte - anomenat *E-Witness* - de *hash* digital per a imatges o vídeo-arxius com empremta digital, ja que és similar l'ús de marques d'aigua amb fotografies, no manipulables, ja que el *hash* original també s'allotja sempre a la cadena de blocs. Aquest *hash* inclouria metadades (lloc, ubicació i algorismes de compressió emprats per fer-ho).

- **Qayyum *et al.* (2019)** en la línia dels autor Hasan i Salah (2019), és a dir, aposta per la tecnologia de *blockchain*, ja que permet la descentralització i aporta transparència, conjuntament amb l'ús de contractes intel·ligents i/o la seva revocació.

La revocació consisteix en la rescissió dels editors de notícies (participants de la cadena de blocs), ja sigui per la seva pròpia sol·licitud o si el sistema ha identificat un determinat editor de notícies que s'ha manifestat "de forma anòmala durant un període de temps específic".

Aquest *smart contract*, establiria com s'inscriuen i es comproven els nous editors de notícies que sol·licitin adherir-se al sistema. A partir d'aquí s'assignarien el parell de claus (pública/privada), verificar aquest editor i assignant-los una puntuació inicial de reputació. Aquest anirà canviant, convertint-se en evolutiva. Aquesta es realitzaria amb el càlcul d'una puntuació de reputació "per a quantificar la credibilitat d'un editor".

Es reforçaria la protecció del material (imatges o vídeos) criptogràficament emmagatzemant-lo en la *blockchain*, ja que "cada interacció amb el contingut és detectable i tractable".

A més a més, el seu plantejament obre les portes a la creació d'un prototip, per a la seva implementació.

- Malgrat una posició majoritària de tots els autors en la utilització de la cadena de blocs per a la detecció de notícies falses i *deepfakes*, per a **Agrawal *et al.* (2021)**, l'enfocament que fan altres autors com és la de buscar l'origen de la *deepfake* (o *fake news* en el seu article) - o identificar com a "potencialment malintencionada" que les pugui publicar – és un enfocament no adequat. Ja que creuen que seria més adequat veure-ho des d'una òptica més de *content curator*, de "seleccionar persones" (reporters o avaluadors) per evitar que es difonguin. Per aquest motiu, Agrawal *et al.* (2021), proposen tres tipologies d'usuaris que interaccionaran amb la *blockchain*:

- Reporters

- Analitzadors

- Validadors

Aquests usuaris adjudicaran una sèrie de punts de credibilitat a la resta per penalitzar-los si propaguen contingut fals, un paradigma similar al plantejat per Qayyum *et al.* (2019) . Per tant un sistema de puntuació que categoritzaria a tots ells. És un model adient per l'anomenat periodisme ciutadà o *citizen journalism* (Barnes, 2012).

- Considerant el treball exposat de Agrawal *et al.* (2021) i Qayyum *et al.* (2019) , veiem similituds amb el que proposen **Jaroucheh *et al.* (2020)**. Els autors, demostren la seva principal inquietud, on la suma de les *fake news*, i el consum de notícies per xarxes socials mitjançant l'efecte *filter-bubbles* - l'efecte bombolla definit per Eli Pariser (2011), que mitjançant els algorismes de cerca aïllen a l'usuari en la seva pròpia bombolla ideològica - deixen a l'usuari al marge, on no té control. Per aquest motiu presenten TRUSTD, un ecosistema digital que vol tornar a posar "a l'usuari al centre de l'equació".

Tres tipologies o entitats que interactuarien amb la cadena de blocs, aprofitarien els factors com la descentralització, l'encriptació i el consens entre les parts:

- El creador de contingut (anomenat CC), que podria ser "en forma de text, foto, vídeo, so o qualsevol altre format digital".

- L'actor avaluador (anomenat AA) o taxador, que obra la porta també a l'anomenat *citizen journalism*.

- l'usuari (o consumidor de notícies).

Els CC escollirien als AA, però l'objectiu d'aquest projecte és que l'usuari tingui el poder d'escollir l'AA dins d'una llista, assignant-los una puntuació o valor de nivell de confiança.

Abans d'entrar en aquest ecosistema, de tipus obert, cada entitat tindria el seu DID, o *Decentralised Identifiers*, identificadors descentralitzats - similars als SSI - per després generar una signatura col·lectiva. TRUSTD hauria de permetre que fos utilitzat per organitzacions, agències de notícies, periodistes, editors, activistes, amb diferents nivells d'experiència i, sobretot pels no-periodistes (s'interpreta com a *street* o *citizen journalism*).

Jaroucheh *et al.* (2020), com Yazdinejad *et al.* (2020), no descarta la integració o ús en la cadena de blocs com a part de la política de confiança entre usuaris, i de programari d'AI, com l'anomenat *Reality Defender*, per detectar fake media.

- Considerant el que han expressat els autors anteriors, **Shae i Tsai (2019)** també volen posar l'èmfasi en una activa participació - parlen de "multitud responsable" - de les persones i no deixar-ho tot en mans de les grans agències de notícies, el govern i els automatismes de l'AI, alhora de filtrar i detectar notícies falses. Per això proposen una plataforma que funcioni com una xarxa social de notícies, on es puguin publicar-les i consultar-les, i pugui ser utilitzades per periodistes i la resta de la població. La plataforma estarà composta per una cadena de blocs amb la integració d'una AI per a la detecció de notícies falses i *deepfake*, ja que les *fake news* "es poden crear a partir de la tecnologia d'IA, però es pot utilitzar la pròpia tecnologia d'IA" per detectar-les.

Aquest ecosistema digital, participarà de l'acció de "consumidors de notícies, creadors de contingut, verificadors de fets de notícies, desenvolupadors de codi de detecció de notícies falses amb AI i editors de mitjans". Per aquest motiu els autors parlen de "la força supervisora comuna de la societat", "esforç interdisciplinari" i de "participació social i política".

Mitjançant unes *News Rooms*, a mode de sales de notícies, de temàtica específica i diferent, és on es podran autenticar els creadors de continguts i periodistes per fer les publicacions i aquestes una vegada completades, seran disponibles per a tothom, i hauran de ser avaluades i classificades dins de la *blockchain*, proporcionant un rànquing.

Aquesta proposta també aprofitarà els avantatges de la *blockchain* en quant a classificació, distribució i traçabilitat, ja que aquesta integració d'AI més cadena de blocs, permetrà saber la "ruta de propagació de notícies".

Gestionat per diferents *smart contracts* i amb un mecanisme de classificació, tindrà un protocol amb dos processos de control:

- Plataforma de distribució (responsable de la confiança dels seus creadors de contingut).

- Plataforma d'edició (responsable de la confiança del seu contingut)

Aquest autors i Jaroucheh *et al.* (2020), també comparteixen similituds amb les inquietuds sobre l'efecte *filter-bubble* ja que Shae i Tsai (2019), afirmen que el seu model pot ser útil "per crear ponts entre diferents comunitats i grups", on en un mateix grup (aïllat) és més fàcil que és difonguin fake news.

- Altrament, **Kietzmann *et al.* (2020)**, dissenyen un marc com a mesura de gestió que podria posar sota control les *deepfakes*, anomenat REAL, acrònim de *Record, Expose, Advocate i Leverage* (enregistrament, exposició, legalitat i estructures confiables).

- En l'àrea de *Record*, creuen fermament en l'ús de tecnologies conjuntes com l'ús d'empremtes digitals - parlen de la marca *Amber Authenticate* - i la cadena de blocs. Una permet introduir l'empremta digital, encriptada, mentre es grava el vídeo o la imatge i la tecnologia, i amb la cadena de blocs permetria la traçabilitat i l'arxivament de manera segura. L'objectiu és obtenir una *alibi service* o *lifelog,* la coartada autenticada per evitar l'efecte del *Liar's dividend* (Chesney i Citron, 2019), una evidencia de l'original, sense manipulacions i autenticada, contra la manipulació o *deepfake*, per a poder revocar-la. Al mateix temps, troben com a desavantatge la tecnologia de rastreig perquè posa al descobert la privacitat de la persona implicada.

- En l'àrea de *Expose*, també advoquen per tecnologies de detecció com les que investiga DARPA.

Aquest dos darrers factors són coincidents amb els plantejats per Yazdinejad *et al.* (2020).

Kietzmann *et al.* (2020) tot i remarcar l'amenaça les *deepfakes* - com també esmenta Mitra *et al.* (2020), quan conceben el *deepfake* com un fenomen dual, ja que podria, per posar un exemple, ajudar persones amb discapacitat auditiva per generar simulació del "moviment dels llavis en acord amb l'àudio" - hi veu algun aspecte positiu i lúdic en la seva pràctica.

- En un altre estudi, **Westerlund (2019)**, aprofita per copsar les inquietuds creixents dels mitjans de comunicació, des d'un enfocament més palpable de la realitat, és a dir, mitjançant articles periodístics que analitzen les tendències del fenomen.

Coincideix, per tant amb altres autors esmentats (Floridi, 2018 citat per Yazdinejad *et al.*, 2020; Hasan i Salah, 2019; i Qayyum *et al.*, 2019), en l'ús de la *blockchain* per a "verificar els orígens i la distribució dels vídeos creant i emmagatzemant signatures digitals en un llibre major gairebé impossible de manipular", en la importància de les eines AI per a la detecció de *deepfakes*; en el pes que han de tenir les empreses de xarxes socials i la coartada autenticada, coincident també amb Kietzmann *et al.* (2020).

- A **Centobelli *et al.* (2021)**, es fa un extens i detallat anàlisi mitjançant l'estudi bibliomètric de 2233 publicacions de WoS (Web Of Science) relacionades amb la cadena de blocs, amb un important creixement de publicacions que arrenca al 2016. Identifica autors, països i acaba fent sis clústers d'interès. D'aquests en destaquem dos, relacionats amb el nostre estudi:

- Clúster 4 (*supply chain applications*) que implica temes com el *big data*, la intel·ligència artificial, el *cloud computing*, el *deep learning*, *forecasting* (models de predicció) i el *machine learning* (ML). Dins

d'aquest subgrup del ML, l'interpreta com un camp per explorar en relació amb la cadena de blocs, posant de relleu el treball de Hasan i Salah (2019), i on es desenvoluparan "nous *hashs* per evitar les falsificacions de continguts digitals"

- Clúster 3 (*security & privacy applications*) detecta la necessitat de desenvolupar nous algoritmes que detectin els *deepfake*s, que es puguin generar a l'entorn digital.

Centobelli *et al.* (2021), també detecta encara algunes problemàtiques relacionades amb la cadena de blocs com són les vinculades amb la privadesa i la seguretat de les dades, i el retorn social que aporta, encara per demostrar.

- **Jing i Murugesan (2021)**, dirigeixen el seu estudi en l'impacte de les *deepfake* en temps de pandèmia, una situació que genera por i ansietat en la població. Aquest article és el que conté els pitjors presagis - respecte les amenaces que comenten altres autors com Alattar *et al.* (2020) o Hasan i Salah (2019) - anunciant que s'està lliurant "una guerra o conflicte híbrid" de ciberatacs, que manipulen i influencien, creant un caldo de cultiu molt negatiu. Amb aquest context, s'hauria de revertir aquest tipus "de corrent tecnològica" abans de que arribi la singularitat tecnològica[20].

Realitzen una revisió profusa de la política de privacitat digital de les dades i proposen "un model d'índex de confiança per evitar *fake news* i *deepfakes*" mitjançant la cadena de blocs, a les xarxes socials. El nivell de credibilitat d'una notícia quedarà definit per aquest índex. Per exemple, "una puntuació baixa en l' índex de confiança voldrà dir que les notícies no són creïbles".

Afirmen que l'intercanvi d'aquestes *deepfake*s i *fake news* en les xarxes socials té un efecte propagador i negatiu, per manca "d'un mecanisme de verificació de notícies eficaç i eficient". Per això amb l'ús de la *blockchain*, el model pot tenir una implementació pràctica per a ser instal·lat al navegador i es pugui utilitzar a les xarxes socials, com Yazdinejad *et al.* (2020) que també plantejava.

[20] una situació o punt de no retorn tecnològic on hi hauria una "explosió d'intel·ligència que deixaria la intel·ligència de l'home enrere" (Good, 1965).

Amb l'ús de la *blockchain* es podrà verificar la font, amb algorisme de consens (un PoT*, Proof of Trust*, per exemple), gaudint d'una privacitat implementant la ZKP (*Zero Knowledge Proof*).

Jing i Murugesan (2021) destaquen algunes oportunitats i fortaleses de la *blockchain* com poden ser:

- Impacte positiu en la societat a l'aplicar les principals característiques de la *blockchain* com poden ser la "confiança, la transparència, la immutabilitat i la descentralització".

- Dades d'usuari protegides (privacitat i seguretat); control de l'usuari als continguts.

- Millora en el mètode de pagament.

- Solució alternativa al *crowdfunding*.

- Verificació de la veritat.

Per contra identifiquen algunes debilitats i amenaces de la *blockchain* com:

- La cadena de blocs a dia d'avui és 'piratejable', degut a la naturalesa de les "construccions criptogràfiques, de l'arquitectura distribuïda i del context" on s'aplica.

- Pot rebre diferents tipologies d'atacs entre d'altres destaquen els atacs del 51%; els atacs als *smart contracts*; atacs a les DNS (sistema de noms de domini, de l'anglès *Domain Name System*); atac DDoS (de denegació de servei, de l'anglès *Distributed Denial-of-service attack*); atacs de privacitat i atacs forquilla o bifurcacions *blockchain*.

Aquestes limitacions actuals de la cadena de blocs poden quedar resoltes amb les futures *quantum blockchain systems*, és a dir, la *blockchain* de computació quàntica.

- **Masood *et al.* (2021)** destaquen el treball de Fraga-Lamas i Fernández-Caramés (2020) i el de Hasan i Salah (2019), com a bons exemples que ajuden a identificar possibles canvis o manipulacions digitals realitzades - des d'una perspectiva de la *digital forensics* - "dins el contingut visual mitjançant l'ús de la cadena de blocs i els contractes intel·ligents".

Per contra, crítica a aquest darrer (Hasan i Salah, 2019) perquè la seva metodologia només pot funcionar correctament i només potser "aplicable si existeixen les metadades dels vídeos" a analitzar.

Assenyalen que la futura investigació en el combat contra la *deepfake* requereix d'un transversalitat i integració de coneixements (física, *deep learning*, AI) per superar *datasets* amb contingut escàs de dades i valors, esbiaixat, "incomplet o sorollós per a les fases d'entrenament" i per tant de detecció.

- En aquesta fase d'entrament i detecció, esmentada en l'anterior paràgraf, **Mitra *et al.* (2020)** presenten el seu estudi com un treball relatiu i complementari al que han fet Hasan i Salah (2019), ja que es pot "aplicar com una eina de detecció de vídeo fals", si es troba per traçabilitat l'origen del vídeo pristí o primigeni.

Aquesta eina que proposen, (que detecta vídeos falsos a partir dels seus fotogrames clau) podria ser útil per a xarxes socials, principalment per detectar *deepfakes*, amb vídeos multimèdia comprimits i versions reduïdes de tamany (d'arxiu), i estan convençuts que també milloraria la detecció - en el model d'entrenament - si s'introduís vídeo amb soroll[21]. Ja que són aquesta tipologia de vídeo els que normalment s'utilitzen per enviar o reenviar.

[21] Definició de *noise* (soroll) a *The Chicago School of Media Theory*, que el defineix com "a senyals aleatoris, imprevisibles i indesitjables, o canvis de senyals, que emmascaren el contingut de la informació". Consultat a: https://lucian.uchicago.edu/blogs/mediatheory/keywords/noise/

L'estudi va utilitzar una base de dades anomenada *FaceForensics++*[22] i part del *dataset* ofert del DFDC (el *Deepfake Detection Challenge*, esmentat per Skibba, 2020) utilitzant 7773 vídeos pristíns y 7765 *deepfakes* de vídeo.

L'èxit d'haver aconseguit altes taxes de detecció - amb "requisits computacionals més baixos" que altres models i sense emprar "enormes quantitats de dades" - podria contradir a Masood *et al.* (2021). Però els mateixos autors, reconeixen que el seu model no ha funcionat com ells esperaven amb els *nozzy videos*, els vídeos amb manca de claredat o definició.

- L'estudi que proposa un altre marc teòric, és el de **Ki Chan *et al.* (2020)**, amb el que es treballa "amb algorismes de detecció i metadades immutables per a verificar la validesa dels continguts digitals", amb cadena de blocs de codi obert de tipus *Hyperledger Fabric*. Per analitzar les parts de vídeo i àudio s'utilitzarà una arquitectura de xarxa neuronal recurrent CNN-LSTM, òptimes per aquest tipus de tasques, alhora que es desenvolupa un *hash* i codificació de la imatge o el vídeo ("característiques representatives úniques"). També es crea una *hash* amb la descripció de la captura de la imatge o el vídeo ("subtítols descriptius del contingut"). Després, aquesta informació ("en un format comprimit, útil per a una ràpida recuperació" o per avaluacions posteriors de procedència) es posa a la cadena de blocs que permetrà el control total sobre la seva identitat al propietari (artista o creador del contingut).

Els autors recorden una de les màximes de la cadena de blocs aplicada en aquest cas: si un usuari no pot rastrejar un contingut fins el seu origen - mitjançant "el contracte intel·ligent associat, cercant els enllaços" amb els que les imatges o els vídeos estan vinculats per parentiu - és que no "serà un contingut confiable". Aquest concepte de "contingut confiable" és exposat amb la mateixa intenció per Hasan i Salah (2019).

Aquest model presenta dos desavantatges com són un de tipus *hardware* (tamany o volum de les transaccions), i l'altre i no menys important, que el

[22] *FaceForensics* ++ "és un conjunt de dades de falsificacions facials que permet als investigadors formar enfocaments basats en l'aprenentatge profund d'una manera supervisada. El conjunt de dades conté manipulacions creades amb quatre mètodes d'última generació, com són *Face2Face, FaceSwap, DeepFakes i NeuralTextures*" (Rössler et al. 2019).

contingut audiovisual o imatge sigui autèntic (o no manipulat) en el punt de recepció.

A més a més, confeccionen una classificació de dues grans tipologies (que recorden a les de Tewari *et al.*, 2020, que hem vist en el marc teòric) de combat contra la *deepfake* en relació amb la cadena de blocs:

- Les mesures preventives, com podrien ser aquelles que garantitzen "la PoA del vídeo mitjançant els registres immutables que permeten el rastreig fins a la font".

- Les mesures reactives o post-reactives, aquelles que apliquen algorismes de detecció mitjançant DL per poder concloure quins continguts han patit modificacions o no.

Considerant el seu treball més aviat de tipus preventiu, reconeixen que té similituds al proposat per Hasan i Salah (2019), però es mostren crítics amb aquest autors, pel fet que la seva proposta no permet la *identity sovereignty* - o identitat sobirana, on el propietari té el control total sobre el seu contingut i de les seves dades, decidint qui pot accedir i com pot accedir - i requereix encriptar tot el model amb la IPFS (amb la informació del Sistema de Fitxers Interplanetari).

- Un altre estudi que treballa amb la IPFS és el de **Alattar et al. (2020)**, en un estudi més detallat en la seva implementació, aposta per una plataforma centralitzada amb PoA, com *Ethereum*, justificant l'ús de la cadena de blocs, per emmagatzemar el vídeo i les dades pel seu anàlisi, enfront d'una base de dades centralitzada tradicional, que encariria el cost. La plataforma centralitzada "permeten només als participants (*autorithy*) de fer les transaccions al llibre major. Es coneixen les identitats d'aquests participants i les seves transaccions es poden auditar en qualsevol moment". Aquesta cadena de blocs, seria específicament una DSP, una plataforma d'emmagatzematge distribuït, de menor cost, per què les bases de dades de vídeo, "necessiten una maquinària d'emmagatzematge extremadament car".

Aquesta DSP és un IPFS, per ser gratuït i popularment utilitzat, i també assenyalen un inconvenients com fan Ki Chan *et al.* (2020), ja que l'ús de la IPFS "comporta el risc de perdre un node [de la xarxa] si aquest node [usuari o servidor] decideix deixar de participar-hi".

La IPFS però, permet emmagatzemar diferents tipologies com poden ser text, àudio, imatge i vídeo, la qual cosa permet fer extensiu el seu prototip, i combatre diferents models de *deepfakes*, com fan Hasan i Salah (2019). A més a més, de permetre la partició d'un arxiu en diferents blocs, aquest s'emmagatzemen en nodes sense duplicació abaratint el cost i mida.

Aquest plantejament es dota també de les marques d'aigua com han fet Yazdinejad *et al.* (2020) i Shweta Jain (2018), però amb dos tipologies: una per a vídeo (incrustat en les vores dels clips de vídeo, i patentada per la històrica companyia de *watermarks Digimarc Corporation*) i una altre per àudio (dins d'un espectre d'àudio imperceptible per l'oïda humana).

Les avantatges de la marca d'aigua és que aporten "un identificador únic que queda vinculat [directament] a la cadena de blocs", i que es poden aplicar depenent la tipologia de *deepfake* (p.e. si una fos de tipus *puppet-master* es podrien aplicar per tot arreu, o si fos de tipus *face-swapping* col·locar-la a la zona del rostre).

Quan s'utilitza una DPS, "obtenim el mateix *hash* per recuperar l'historial i la procedència dels arxius de vídeo", per poder rastrejar-los. Les metadades del vídeo, però "s'emmagatzemen en una xarxa de cadena de blocs separada". Per aquest motiu, verificació o autentificació d'aquestes marques d'aigua es la realitzaria en dues etapes, una d'elles aprofitant l'accés a les metadades dins de la cadena de blocs.

A més a més, s'utilitzarien dos codis d'identificació:

- un número identificador de contingut (CID), per a identificar i recuperar l'arxiu complet. El CID ha sigut generat per encriptació de tipologia SHA2-256. Qualsevol modificació o manipulació del vídeo, variaria com s'ha demostrat, el seu *hash*.

- un número d'identificació de vídeo (VIN), inclosa en la marca d'aigua, que establiria la correspondència un a un amb el nombre de bloc al *blockchain*.

- Per finalitzar, esmenten el treball més referenciat pels altres articles analitzats, que és la proposta de **Hasan i Salah (2019)**, que utilitzen "els contractes intel·ligents de *Ethereum* per localitzar i rastrejar la procedència i l'historial del contingut digital fins a la seva font original". La font original pot ser de qualsevol naturalesa de contingut digital com "àudios, fotografies, imatges o text".

Com s'ha definit en el marc teòric, els autors emprenen el contracte intel·ligent de la cadena de blocs, com a pedra de toc del seu projecte, que crear el vincle (i la traçabilitat), amb cada vídeo i mantenir una relació jeràrquica (pare/fill) amb les possibles (i autoritzades) obres derivades del vídeo pristí. En aquest cas parlen d'artista original (OA, com a creador o propietari) i d'artista secundari (SA, aquell que vulgui fer-ne una obra derivada) .

Aquest *smart contract* "utilitzarà els *hash* del sistema d'arxius interplanetari (IPFS) que s'utilitzarà per a emmagatzemar els arxius digitals i el contingut" conjuntament amb les metadades. És a dir el vídeo queda emmagatzemat, de manera distribuïda, a la IPFS i el *hash* als *smart contracts*. Per aquest motiu, per la característica immutabilitat i integritat de la cadena de blocs, qualsevol canvi al contingut digital, ja sigui un vídeo o altri, "donarà lloc a la creació d'un *hash* que no coincidirà amb el *hash* que ha quedat registrat al contracte intel·ligent".

Per demostrar l'aposta per la cadena de blocs, fan un paral·lelisme entre el món real i el món digital quan comparen certificats d'autenticitat (CoA) - on aquests són manipulables o falsificables - i les PoA (la prova d'autenticitat), eina cabdal per evitar-ho. Mostren altres avantatges característiques de la *blockchain* com són la distribució descentralitzada; o de com les dades i transaccions queden registrades ("les dades de procedència son accessibles, i disponibles ja que són replicades a tots els nodes de la xarxa") en el *ledger* o llibre mestre; l'ús de la clau privada o la generació dels *timestamps*.

També fan un valoració de costos, ja que "per a cada transacció realitzada a la cadena de blocs [al utilitzar una plataforma Ethereum] hi ha un cost que s'expressa en termes de Gas i *tokens Ether* (ETH)".

El seu treball és aplicable també per a DApps (*decentralized applications)* o aplicacions descentralitzades (*Twitter* seria un exemple contrari, ja que és una aplicació centralitzada) o com extensió per a navegadors, com el treball que proposa Yazdinejad *et al.* (2020).

Esmenten el treball específic empreses emergents com *Prover* i *OriginalMy*, o altres estudis acadèmics com el de Gipp *et al.* (2016) que utilitzen la *blockchain* només enfocada en una part: o bé per utilitzar el *hash* mentre l'usuari està capturant el vídeo o bé només enfocats només en la integritat del vídeo.

6. CONCLUSIONS

Després d'haver fet la cerca en diferents repositoris per tal de respondre a la hipòtesi plantejada inicialment, i haver fet una selecció seguint un criteri de rellevància i impacte es pot arribar a diverses conclusions.

Responent a la hipòtesi formulada, la metodologia *blockchain* (cadena de blocs), i els elements que la componen, ens aporta diferents aproximacions per solucionar els problemes d'autenticació de les imatges en el context actual del fenomen *deepfake* (vídeo hipertrucat), com poden ser:

- l'ús dels contractes intel·ligents per establir la relació dels usuaris amb la cadena de blocs, també per a definir la relació jeràrquica entre ells i l'allotjament del *hash* de vídeo.

- la importància de triar quins protocols i xarxes P2P, o tipologies de *blockchain* és més adient. Actualment les propostes s'han centrat en l'ús de la IPFS per al vídeo, i emprar entorns com *Etherum* (*tokens* en ETH) o *Hyperledger Fabric* (registre no públic, codi obert).

La comunitat acadèmica i científica estudiada, veu 'una guerra freda' o 'un joc del gat i la rata' entre la creació de *deepfake* i els sistemes per detectar-la. I han demostrat que les característiques pròpies d'una cadena de blocs com són (les conegudes en anglès com "les *4T*") traçabilitat, seguiment, transparència i confiança, són adients per adaptar-les al camp d'estudi plantejat. Però és necessària la combinació amb altres elements:

- L'ús d'eines AI de detecció, com a pas previ o procedimental, que asseguri que el contingut digital del vídeo pristí (imatge, àudio, etc...) és autèntic.

- Les marques d'aigua (*watermarks*) d'àudio, vídeo i imatge, durant l'enregistrament amb l'ús del *hash*.

- Altres metadades en el moment de l'enregistrament, també en *hash values* (empremtes digitals de valor numèric). Que permetrien "la coartada autenticada" o *lifelogging*.

- El compromís i implicació de les xarxes socials amb la societat, per reduir-ne l'impacte, fomentant la fiabilitat.

- L'establiment d'un "carnet per punts" de credibilitat i fiabilitat en relació amb els usuaris, organismes i contingut digital.

- Definir el rol de l'usuari, qui hi participa i com hi participa.

- Proactivitat del ciutadà.

6.1. Futures línies d'estudi

A la majoria d'estudis s'exposa un marc teòric, conceptual i prototípic, es pot plantejar com a línia futura d'investigació fer un seguiment de la implementació real i verídica del *blockchain* en aquesta àrea tant concreta que es desenvolupa en aquest treball, i que pugui donar resposta al que suposa l'ús de la cadena de blocs com són les mancances de privacitat (rastreig fins a l'usuari) i de seguretat (tipologies d'atacs i debilitat en l'encriptació), i comprovar el ROI o retorn d'inversió.

A més, tenint en compte que el *blockchain* és un tecnologia disruptiva que ja s'aplica al dia a dia i ho serà encara més en un futur proper, seria interessant desenvolupar un sistema didàctic i pedagògic que apropi el concepte a la societat degut a la seva complexitat. L'eina més efectiva per fer-ho, a curt termini, seria mitjançant l'alfabetització mediàtica i informativa (*Media & Information Literacy*).

Proposem que es segueixin investigant altres eines, com l'ús dels CBIR o reaprofitar sistemes similars que ja disposem com el LOCKSS, que empra el CSUC, una eina automatitzada que compara periòdicament les còpies dels arxius.

7. BIBLIOGRAFIA

Ab Rahman, A., Hamid, U. Z. A., i Chin, T. A. (2017). Emerging technologies with disruptive effects: a review. Perintis eJournal, 7(2), 111-128 https://bit.ly/3fYdpXy

Agirreazaldegi Berriozabal, T. (2011). La gestión de materiales audiovisuales de programas informativos en las cadenas de televisión generalistas. *BiD: textos universitaris de biblioteconomia i documentació*, núm. 26 (juny). https://dx.doi.org/10.1344/105.000001729

Agrawal, P., Singh Anjana, P., & Peri, S. (2021). DeHiDe: Deep Learning-based Hybrid Model to Detect Fake News using Blockchain, *International Conference on Distributed Computing and Networking*. Association for Computing Machinery, New York, NY, USA, 245–246 https://doi.org/10.1145/3427796.3430003

Alattar, A., Sharma, R., i Scriven, J. (2020). A System for Mitigating the Problem of Deepfake News Videos Using Watermarking. *Electronic Imaging*, 2020(4), 117-1. https://doi.org/10.2352/ISSN.2470-1173.2020.4.MWSF-117

Albahar, M., & Almalki, J. (2019). Deepfakes: Threats and Countermeasures Systematic Review. *Journal of Theoretical and Applied Information Technology*, 97(22), 3242-3250. https://bit.ly/34W2HdR

Barnes, C. (2012). Citizen journalism vs. traditional journalism: A case for collaboration. *Caribbean Quarterly*, 58(2), 16-27,179. https://bit.ly/3z80VUQ

Bregler, C., Covell, M., Slaney, M. (1997). Video rewrite: driving visual speech with audio. *SIGGRAPH'97: Proceedings of the 24th annual conference on computer graphics and interactive techniques*, p. 353–360.

Brigden, N. (Director, amb Hany Farid). (2020). Chapter 4: Digits. *Connected: The Hidden Science of Everything*. [Documental]. Netflix.

Bui, T., Cooper, D., Collomosse, J., Bell, M., Green, A., Sheridan, J., ... & Brown, A. (2019). Archangel: Tamper-proofing video archives using temporal content hashes on the blockchain. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*

Canton Ferrer, C., Dolhansky, B., Pflaum, ... Lu, J (2020). Deepfake Detection Challenge Results. *ML Applications Computer Vision*. Facebook AI. https://bit.ly/3gg1KlY

Casino, F., Dasaklis, T.K. i Patsakis, C. (2019). A systematic literature review of blockchain-based applications: Current status, classification and open issues, Telematics and Informatics, Volume 36, 2019, pp. 55-81, https://doi.org/10.1016/j.tele.2018.11.006

Centobelli, P., Cerchione, R., Esposito, E., & Oropallo, E. (2021). Surfing blockchain wave, or drowning? Shaping the future of distributed ledgers and decentralized technologies, *Technological Forecasting and Social Change*, Volume 165 https://doi.org/10.1016/j.techfore.2020.120463

Chaum, D. L. (1981). Untraceable electronic mail, return addresses, and digital pseudonyms. Commun. ACM 24, 2, 84–90. https://doi.org/10.1145/358549.35

Chesney, B., i Citron, D. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs*. https://fam.ag/3v0mrri

Chesney, B., i Citron, D. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. Calif. L. Rev., 107, 1753

Cherry, K., i Lustik C.(2020). What Is the Uncanny Valley?. VeryWell Mind. https://bit.ly/3in5bKj

Corrigan, J. (2021, 13 Abril). DARPA Is Taking on the Deepfake Problem. Nextgov.Com. https://bit.ly/3v174z5

Dai, W. (1998). B-Money-an anonymous, distributed electronic cash system. https://bit.ly/3cpj46G

Engler, A. (2020). Fighting deepfakes when detection fails. *Brookings*. https://brook.gs/3v3R5A2

Farid, H. (2021). On Algorithmic Amplification. *Inference: International Review of Science*. https://bit.ly/2S7YMIh

Fraga-Lamas, P. & Fernández-Caramés, T. M. (2020). Fake News, Disinformation, and Deepfakes: Leveraging Distributed Ledger Technologies and Blockchain to Combat Digital Deception and Counterfeit Reality, *IT Professional*, vol. 22, no. 2, pp. 53-59 https://doi.org/10.1109/MITP.2020.2977589

Frankenfield, J. i Rasure, E. (2021). Distributed Ledger Technology (DLT). *Investopedia*
https://bit.ly/3g9ESUY

García-Morales, E. (2018). "Luces y sombras sobre el impacto del blockchain en la gestión de documentos". Anuario ThinkEPI, v. 12, pp. 345-351.
https://doi.org/10.3145/thinkepi.2018.58

Giancaspro, M. (2017). Is a 'smart contract' really a smart idea? Insights from a legal perspective. *Computer Law & Security Review*, Volume 33, Issue 6, Pages 825-835,
https://doi.org/10.1016/j.clsr.2017.05.007

Gipp, B., Kosti, J. i C. Breitinger, C. (2016). Securing video integrity using decentralized trusted timestamping on the bitcoin blockchain, *Proc. MCIS,* p. 51

Good, I. J. (1965). Speculations Concerning the First Ultraintelligent Machine. *Accelerating Future*.
https://bit.ly/3w5wi0y

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial networks.
https://arxiv.org/abs/1406.2661

Grech, A. i Camilleri, A.F. (2017). Blockchain in Education. Joint Research Centre (JRC), European Union

Hasan, H. R. i Salah, K. (2019). Combating Deepfake Videos Using Blockchain and Smart Contracts, *IEEE Access*, vol. 7, pp. 41596-41606
https://doi.org/10.1109/ACCESS.2019.2905689

Hausser, D. Secure decentralized pricing and accounting for peer-to-peer systems.
https://doi.org/10.3929/ethz-a-005163959

Hellman, M. (1976). New directions in cryptography. IEEE transactions on Information Theory, 22(6), 644-654.
https://doi.org/10.1109/TIT.1976.1055638

Hongling, L., i Di, W. (2019). Application of Asymmetric Key Technology in M-ES, 2018 11th International Conference on Intelligent Computation Technology and Automation (ICICTA), pp. 186-189,
https://doi.org/10.1109/ICICTA.2018.00049

Houben, R. i Snyersc, A. (2018). Cryptocurrencies and blockchain: Legal context and implications for financial crime, money laundering and tax evasion, *European Union*,
https://bit.ly/3gg0grP

I2CAT (2020). Informe de tendències sobre la tecnologia blockchain, *Departament de Polítiques Digitals*, Generalitat de Catalunya.

Jaroucheh, Z., Alissa, M., Buchanan, W. J. and Liu, X. (2020). TRUSTD: Combat Fake Content using Blockchain and Collective Signature Technologies, *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2020, pp. 1235-1240 https://doi.org/10.1109/COMPSAC48688.2020.00-87

Jing T. W., Murugesan., R. K. (2021). Protecting Data Privacy and Prevent Fake News and Deepfakes in Social Media via Blockchain Technology, *Advances in Cyber Security*. Communications in Computer and Information Science, vol. 1347 https://doi.org/10.1007/978-981-33-6835-4_44

Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2019). Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 8110-8119).

Kaplan, F. (2015). The Venice Time Machine. *In Proceedings of the 2015 ACM Symposium on Document Engineering (DocEng '15)*. Association for Computing Machinery, New York, NY, USA, 73. https://doi.org/10.1145/2682571.2797071

Ki Chan, C. C., Kumar V., Delaney, S. & Gochoo, M. (2020). Combating Deepfakes: Multi-LSTM and Blockchain as Proof of Authenticity for Digital Media, *2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G)*, pp. 55-62 https://doi.org/10.1109/AI4G50087.2020.9311067

Kietzmann, J., Lee, L. W., McCarthy, I. P. & Kietzmann, T. C. (2020). Deepfakes: Trick or treat?, *Business Horizons*, Volume 63, Issue 2, pp. 135-146, https://doi.org/10.1016/j.bushor.2019.11.006

Korshunov, P., i Marcel, S. (2018). Deepfakes: a new threat to face recognition? assessment and detection.

Marcet, X. (2018). Els arxius en temps del Blockchain. Arxivers la DaDa. https://arxivers.com/ladada/els-arxius-en-temps-del-blockchain/

Marcet, X. (2020). Papers de l'Observatori de la Indústria, *Departament d'Empresa i Coneixement de la Generalitat de Catalunya* https://bit.ly/3x1kqfW

Marques, O., i Furht, B. (2002). MUSE: A Content-Based Image Search and Retrieval System Using Relevance Feedback. *Multimedia Tools Appl.* 17. 21-50

Malik, Y. S., Sabahat N., and Moazzam, M. O. (2020). Image Animations on Driving Videos with DeepFakes and Detecting DeepFakes Generated Animations, *2020 IEEE 23rd International Multitopic Conference (INMIC)*, pp. 1-6 https://doi.org/10.1109/INMIC50486.2020.9318064

Masood, M., Nawaz, M., Malik, K. M., Ali Javed, A., & Irtaza, A. (2021). Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward. arXiv preprint. Cornell University https://arxiv.org/abs/2103.00484

Marqués-Pascual, J. (2020). Tecnologia blockchain: cap a un canvi disruptiu en la informació. Marqués-Pascual, J., Sintes Olivella, M. (coord.) Blockchain i periodisme: com la cadena de blocs canviarà els mèdia. Barcelona, Editorial UOC

Microsoft Corporation. (2018). JFK Files, *Articial Intelligence* https://bit.ly/2T87LsQ

MIT Open Learning. (2020). Tackling the misinformation epidemic with "In Event of Moon Disaster". *MIT News | Massachusetts Institute of Technology*. https://bit.ly/3v5nq9z

Mitra, A., Mohanty, S. P., Corcoran, P. *et al.* (2021). A Machine Learning Based Approach for Deepfake Detection in Social Media Through Key Video Frame Extraction. *SN COMPUT. SCI.* 2, 98 https://doi.org/10.1007/s42979-021-00495-x

Moore, D., i Rid, T. (2016). Cryptopolitik and the Darknet. Survival, 58(1), 7-38. https://doi.org/10.1080/00396338.2016.1142085

Nakamoto, S., (2008). Bitcoin: A peer-to-peer electronic cash system. https://bit.ly/3ip230r

Newman, N., Fletcher, R., Schulz, A., et al.(2020). Digital News Report 2020, *Reuters Institute for the Study of Journalism* https://bit.ly/350caAZ

Octavian Report. (amb Hany Farid, 2019). Impostor Syndrome. https://bit.ly/2TbwUTO

Panetta, K. (2020). Gartner Top 10 Strategic Technology Trends for 2020, Smarter With Gartner. https://gtnr.it/2S8IGy7

Pariser, E. (2011). The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think. Penguin Books.

Pournader, M., Shi, Y., Seuring, S. i Lenny Koh, S. C. (2020). Blockchain applications in supply chains, transport and logistics: a systematic review of the literature, *International Journal of Production Research*, https://doi.org/10.1080/00207543.2019.1650976

Qayyum, A., Qadir, J., Janjua., M. U. & Sher, F. (2019). Using Blockchain to Rein in the New Post-Truth World and Check the Spread of Fake News, *IT Professional*, vol. 21, no. 4, pp. 16-24 https://doi.org/10.1109/MITP.2019.2910503

Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Nießner, M. (2019). Faceforensics++: Learning to detect manipulated facial images. *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1-11).

Ruipérez García, R. (2019). Blockchain o Cadena de bloques / Entrevistat per Isabel Baeza. *El reto de la tecnología Blockchain*. Canal UNED. https://canal.uned.es/video/5cd1331da3eeb0560f8b456a

Samanta, P., i Jain, S. (2018). E-Witness: preserve and prove forensic soundness of digital evidence. Proceedings of the 24th Annual International Conference on Mobile Computing and Networking https://doi.org/10.1145/3241539.3267720

Sarrias, S. (2020). Del bitcoin a la blockchain: aplicaciones para el sector público, Sessió sobre Blockchain (IMI), La Salle-URL, Barcelona

Serra, J. (2020). Tinc set de dades!, *Arxivers la DaDa*. https://arxivers.com/ladada/tinc-set-de-dades/

Sorkin, A. D. (2017). Close Read: In Event of Moon Disaster. *The New Yorker*. https://bit.ly/3gdX9AJ

Shae, Z. & Tsai, J. (2019). AI Blockchain Platform for Trusting News, *2019 IEEE 39th International Conference on Distributed Computing Systems* https://doi.org/10.1109/ICDCS.2019.00160

Skibba, R. (2020). Accuracy Eludes Competitors in Facebook Deepfake Detection Challenge, Engineering, Volume 6, Issue 12, pp.1339-1340 https://doi.org/10.1016/j.eng.2020.10.008

Solomon, M.G. (2019). Enterprise Blockchain: Oracle Special Edition, For Dummies. Wiley Brand, https://bit.ly/3zc5bT8

Suwajanakorn, S., Seitz, S. M., i Kemelmacher-Shlizerman, I. (2017). Synthesizing obama: learning lip sync from audio. ACM Transactions on Graphics (ToG), 36(4), 1-13.

Szabo, N. (1997). Formalizing and Securing Relationships on Public Networks. First Monday, 2(9). https://doi.org/10.5210/fm.v2i9.548

Tapscott, D., i Tapscott, A. (amb Dans, E.). (2017). La revolución blockchain. Deusto.

Tewari, A., Fried, O., Thies, J., Sitzmann, V., Lombardi, S., Sunkavalli, K., … Nießner, M. (2020). *State of the Art on Neural Rendering*. *Computer Graphics Fo*rum, 39(2), 701–727.

The Economist. (2015). *The trust machine*. https://econ.st/3ionI8Z

Thies, J., Zollhofer, M., Stamminger, M., Theobalt, C., & Nießner, M. (2016). Face2face: Real-time face capture and reenactment of rgb videos. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2387-2395).

Ufarte Ruiz, M. J. (2012). La situación laboral del periodista como factor condicionante de la calidad informativa: con precariedad no hay calidad. *Actas IV Congreso Internacional Latina de Comunicación Social: comunicación, control y resistencias*. https://dialnet.unirioja.es/servlet/articulo?codigo=4222723#?

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review,* 9(11), 39-52. http://doi.org/10.22215/timreview/1282

Yazdinejad A., Parizi, R. M., Srivastava, G., & Dehghantanha, A. (2020). *Making sense of blockchain for AI deepfakes technology*. IEEE Globecom Workshops pp.1-6. http://dx.doi.org/10.1109/GCWkshps50303.2020.9367545

Zamorano, V. (2018). El problema de los generales bizantinos. *Blockchain Services*. https://bit.ly/3gfg5Pp