

CAÇADORS D'HIPERTRUCATGES. REPTES I SOLUCIONS TECNOLÒGIQUES PER A LA DETECCIÓ I AUTENTICACIÓ*

Sergio Colado García

Professor de la Universitat Internacional de La Rioja (UNIR) i director del Nechi Group (Barcelona). www.sergiocolado.com. scolado@nechigroup.com

Resum: En una societat cada vegada més digitalitzada, els hipertrucatges, tecnologies basades en intel·ligència artificial, representen un repte creixent. Aquests continguts multimèdia falsos poden imitar imatges, vídeos i àudios amb gran realisme, i generar preocupacions sobre la seva utilització per a la manipulació i la desinformació. Aquest article explora què són els *hipertrucatges*, les tècniques utilitzades per crear-los i els mètodes avançats de detecció, com les xarxes neuronals convolucionals (en anglès, *convolutional neural networks*, CNN) i la reconfiguració de xarxes generatives adversàries (en anglès, *generative adversarial network*, GAN). També s'aborda la importància de l'autenticació del contingut digital mitjançant tècniques com les signatures digitals i la cadena de custòdia. L'article conclou amb una reflexió sobre la necessitat d'una resposta coordinada, tant tecnològica com legal, per mitigar els riscos associats als hipertrucatges.

Paraules clau: hipertrucatges, intel·ligència artificial, detecció de manipulació, autenticació digital, xarxes neuronals.

DEEPPFAKE HUNTERS. CHALLENGES AND TECHNOLOGICAL SOLUTIONS FOR DETECTION AND AUTHENTICATION

Abstract: In an increasingly digitalized society, deepfakes, which are AI-based technologies, represent a growing challenge. These false multimedia contents can realistically mimic images, videos, and audio, raising concerns about their potential use for manipulation and disinformation. This article explores what deepfakes are, the techniques used to create them, and advanced detection methods, such as convolutional neural networks (CNNs) and generative adversarial network (GAN) reconfiguration. It also addresses the importance of digital content authentication through techniques like digital signatures and chain of custody. The article concludes with a reflection on the need for a coordinated response, both technological and legal, to mitigate the risks associated with deepfakes.

Keywords: deepfakes, artificial intelligence, manipulation detection, digital authentication, neural networks.

Introducció

En una societat cada vegada més digitalitzada, distingir entre el que és real i el que és fals esdevé un desafiament. Els avenços en intel·ligència artificial han donat lloc als hipertrucatges (*deepfakes*), tecnologies capaces de generar imatges, vídeos i àudios que són gairebé indistingibles dels autèntics. Aquests continguts, que combinen l'aprenentatge profund (*deep learning*) amb informació falsa (*fakes*), poden tenir usos creatius i lú-

dics, però també poden ser eines poderoses per crear desinformació i manipulació massiva.

Els hipertrucatges representen un desafiament tecnològic i social profund en una societat altament digitalitzada, com s'ha explorat en treballs previs sobre tecnologia digital (Colado, 2021), ja que poden afectar la política, l'economia i la vida personal. Amb aquesta amenaça en augment, les eines d'intel·ligència artificial (IA) no només es veuen com una part del problema, sinó també com una solució clau per detectar-los i controlar-los.

Què és un hipertrucatge?

Segons el TERMCAT (2024), el terme correcte per a *deepfakes* en català és *hipertrucatge*. Un hipertrucatge és un contingut multimèdia fals creat mitjançant tècniques d'aprenentatge automàtic i xarxes neuronals profundes. S'utilitzen dues tècniques principals per generar-los: les

* Aquest article s'ha elaborat a partir de la revisió, ampliació i actualització d'un treball anterior publicat per l'autor sota el títol «Deepfake Hunters. Caçadors de mentides» al lloc web personal sergiocolado.com el 26 de juny de 2024. Tot i que el text original plantejava una introducció al fenomen dels hipertrucatges, aquesta versió ha estat reestructurada per aprofundir en els aspectes tècnics, ètics i socials, i integra exemples nous, desenvolupaments tecnològics recents i suggeriments pràctics per a la detecció i autenticació de continguts digitals. A més, s'ha adaptat per incloure la terminologia catalana normalitzada pel TERMCAT.

xarxes generatives adversàries (en anglès, *generative adversarial network*, GAN) i els autoencoders.

Les GAN consisteixen en un model de dues xarxes neuronals que treballen de manera competitiva. Per una banda, un generador crea continguts falsos, com ara imatges o vídeos. Per l'altra, un discriminador intenta distingir entre el contingut fals i el real per aprendre'l. Aquesta competició fa que el generador millori constantment fins a crear continguts gairebé indistingibles dels originals.

Els autoencoders són models entrenats per reconstruir dades d'entrada (com una cara o una veu) i després modificar-les. Es poden utilitzar per transferir l'expressió facial d'una persona a una altra, generant imatges o vídeos en què sembla que el subjecte fa moviments o diu paraules inexistents. Els vídeos en què una persona famosa sembla dir o fer coses que mai no ha fet en són un exemple.

La creació d'un hipertrucatge implica l'ús de grans quantitats de dades de vídeos o imatges del subjecte que s'ha de falsificar, un procés de preentrenament d'un model basat en IA i l'ús de tècniques com la superresolució per millorar-ne el realisme. Una vegada entrenat, el model pot produir vídeos, àudios o imatges que simulen fidelment el subjecte real. Aquesta semblança amb la realitat accentua el risc del seu ús malintencionat.



FIGURA 1. Exemple d'hipertrucatge. Superposició de cares.
FONT: Elaboració pròpia.

Hi ha diversos usos dels hipertrucatges a la indústria. Per exemple, en cinema i entreteniment es poden recrear personatges històrics o millorar efectes especials. En educació i formació, es poden simular escenaris realistes o crear assistents virtuals. I en medicina permet estudiar comportaments humans o millorar simulacions quirúrgiques.

Però també hi ha usos molt negatius, com ara la creació de vídeos falsos de líders polítics dient o fent coses comprometedores, l'elaboració de contingut fals que danya la reputació de les persones o la simulació de la veu d'un directiu per autoritzar transferències il·legítimes, entre d'altres.

El potencial d'aquesta tecnologia de confondre entre allò real i fals planteja reptes importants que cal analitzar amb profunditat. Amb l'evolució dels hipertrucatges, el desenvolupament d'eines per a la seva detecció i autenticació és essencial per minimitzar-ne els riscos i garantir un ús ètic d'aquesta tecnologia.

A la caça dels hipertrucatges

Amb l'evolució dels hipertrucatges, detectar continguts manipulats s'ha convertit en un desafiament que exigeix l'ús combinat de diverses eines tecnològiques. La millora contínua dels algorismes d'intel·ligència artificial, juntament amb les eines d'anàlisi forense digital, ha fet possible identificar anomalies que són imperceptibles a simple vista.

Els primers hipertrucatges presentaven errors fàcilment detectables, com la desincronització labial o anomalies en les textures facials. No obstant això, amb l'avenç tecnològic, aquests errors s'han anat minimitzant, i avui en dia els hipertrucatges més sofisticats poden imitar amb gran precisió expressions facials, moviments corporals i patrons vocals, cosa que fa molt més difícil detectar-los a simple vista.



FIGURA 2. Recreació d'un caçador d'hipertrucatges.
FONT: Elaboració pròpia amb IA.

Per combatre aquesta amenaça en evolució constant, s'han desenvolupat eines tecnològiques avançades, especialment basades en intel·ligència artificial. A continuació es detallen els principals mètodes utilitzats:

1. Anàlisi de consistència visual

Els hipertrucatges poden contenir anomalies subtils que, tot i no ser evidents a primera vista, es poden detectar amb tècniques automatitzades:

- Textura de la pell: sovint presenten incoherències en la textura de la pell, que pot semblar artificial o uniforme en algunes zones.

- Il·luminació i ombres: els vídeos generats poden presentar discrepàncies en la il·luminació o en la posició de les ombres, que no són coherents amb les fonts de llum del vídeo original.

- Geometria facial: l'anàlisi de les proporcions facials i la seva coherència al llarg de diferents fotogrames permet detectar alteracions subtils que afecten la morfologia del rostre.

2. Anàlisi de freqüència

L'anàlisi de freqüències també ha demostrat ser una eina clau per detectar artefactes generats durant el procés de

manipulació, mitjançant tècniques com la transformada de Fourier (Gong, Li i Wang, 2020). Aquesta tècnica consisteix a descompondre les imatges o els vídeos en els seus components de freqüència, la qual cosa permet identificar artefactes generats durant el procés de creació de l'hipertrucatge:

- Transformada de Fourier: permet analitzar la freqüència de les imatges per detectar-hi patrons irregulars o artefactes ocults.

- Espectres de freqüència: els hipertrucatges poden deixar empremtes específiques en certes bandes de freqüència que revelen la manipulació.

3. Detecció d'artefactes digitals

Els hipertrucatges solen generar petits artefactes que poden ser detectats amb tècniques forenses:

- Comprensió d'imatges i vídeos: les imatges i els vídeos falsificats poden presentar patrons de comprensió diferents dels originals, la qual cosa crea artefactes visibles en les zones manipulades.

- Errors en la codificació de píxels: les inconsistències en la disposició dels píxels o en la representació del color poden ser indicadors de manipulació digital.

4. Anàlisi biomètrica i microexpressions

La detecció de petites anomalies en els moviments facials i corporals pot ser una manera efectiva de descobrir hipertrucatges:

- Moviments facials: les microexpressions, que són moviments ràpids i involuntaris del rostre, són difícils de replicar de manera precisa. L'anàlisi d'aquests petits gestos pot revelar inconsistències en el vídeo.

- Sincronització labial: la desincronització entre el moviment dels llavis i l'àudio, especialment en vídeos amb diàlegs, és un senyal comú de manipulació.

5. Anàlisi d'àudio i veu

A més del contingut visual, l'àudio també pot oferir pistes importants sobre la falsificació:

- Freqüència i timbre de la veu: els hipertrucatges poden generar distorsions subtils en la veu, especialment en la freqüència i el timbre, que poden ser identificades amb anàlisi acústica.

- Inconsistències lingüístiques: l'estil de parla, l'entonació o l'ús de determinades paraules poden no coincidir amb el patró habitual de la persona, la qual cosa en revela una falsificació.

6. Mètodes basats en xarxes neuronals

Les tècniques d'aprenentatge automàtic també s'utilitzen per detectar contingut manipulat:

- Xarxes neuronals convolucionals (CNN): aquestes xarxes estan entrenades per detectar petites anomalies en les imatges, com errors en la textura o en la geometria facial (Shad, López i Kumar, 2021).

- Xarxes neuronals recurrents (RNN): aquestes xarxes analitzen seqüències de vídeo per detectar incoherències temporals en el moviment dels personatges, identificant errors subtils en la fluïdesa dels gestos i expressions.

cies temporals en el moviment dels personatges, identificant errors subtils en la fluïdesa dels gestos i expressions.

- GAN reconfigurades: tot i que les xarxes generatives adversàries (GAN) s'utilitzen per crear hipertrucatges, també es poden reconfigurar per detectar anomalies generades per altres GAN (Gong, Li i Wang, 2020).

7. Anàlisi de metadades i estructura del fitxer

Els hipertrucatges no només alteren el contingut visual, sinó que també poden introduir incoherències en la informació associada al fitxer:

- Metadades alterades: l'anàlisi de les metadades pot revelar discrepàncies en les dates de creació o en la informació del dispositiu de gravació, cosa que n'indica una possible manipulació.

- Anomalies en l'estructura del fitxer: canvis inesperats en la codificació o en la comprensió del fitxer poden ser signes d'alteració.

Aquest conjunt de mètodes de detecció representa una eina poderosa per identificar hipertrucatges, i la seva efectivitat depèn de la seva aplicació conjunta i de l'evolució constant de les tècniques utilitzades per crear i detectar aquests continguts falsificats (Shad, López i Kumar, 2021).

L'ús de tots aquests mètodes, combinat amb el desenvolupament continu de noves eines tecnològiques, permet avançar en la detecció proactiva d'hipertrucatges. La lluita contra aquesta tecnologia evolutiva exigeix una millora constant de les eines forenses i l'ús de tècniques d'intel·ligència artificial capaces de seguir el ritme de les innovacions en la creació de continguts falsificats. Només mitjançant una estratègia multidisciplinària i tecnològicament avançada es pot garantir una detecció efectiva d'aquests continguts.

Autenticació del contingut digital

L'autenticació del contingut digital, especialment en l'era dels hipertrucatges, és un procés essencial per garantir la integritat i autenticitat de vídeos, imatges i àudios, sobretot quan es presenten com a proves en tribunals o en investigacions periodístiques. A mesura que les tècniques de manipulació digital esdevenen més sofisticades, és imprescindible utilitzar tècniques avançades que permetin determinar si un fitxer digital ha estat alterat.

El procés d'autenticació comença amb la cadena de custòdia, que és un element fonamental per mantenir la credibilitat de qualsevol prova digital. Aquesta cadena requereix un registre detallat i ininterromput del contingut des del moment en què es crea o es recull fins que es presenta. Això implica assegurar que el contingut ha estat emmagatzemat i gestionat adequadament, sense cap accés no autoritzat ni manipulació, la qual cosa en garanteix la integritat. Als tribunals, sovint es rebutgen proves digitals si no es pot demostrar una cadena de custòdia clara.

Les metadades són una altra eina clau per a l'autenticació. Inclouen informació associada a un fitxer digital, com ara la data de creació, el tipus de dispositiu utilitzat, la ubicació geogràfica (si s'ha activat el GPS) i els canvis o modificacions anteriors. Les metadades aporten informació essencial sobre l'origen i la història del contingut. L'anàlisi forense de les metadades pot revelar anomalies, com ara dates incoherents o localitzacions inexactes, que podrien indicar que el fitxer ha estat manipulat.

A més de les metadades, s'utilitzen tècniques de marcatge o marques d'aigua digitals, que consisteixen a incrustar informació oculta o visible en imatges o vídeos. Aquestes marques poden incloure detalls sobre l'autor o el propietari, o bé servir com a prova d'autenticitat. Les marques d'aigua són especialment útils perquè poden ser invisibles a l'usuari, però detectables a través de programari especialitzat, i serveixen com a defensa davant possibles manipulacions. Aquest mètode s'utilitza sovint en arxius multimèdia d'alta importància o amb informació confidencial, que cal garantir que no hagin estat alterats.

Un altre mecanisme important és l'ús de signatures digitals. A través de tècniques criptogràfiques, una signatura digital s'associa a un fitxer per garantir-ne la integritat i confirmar-ne l'autenticitat. Aquest procés implica la creació d'una empremta electrònica única del fitxer en el moment de la creació o verificació inicial, que actua com una «empremta digital» del contingut. Si el contingut és alterat, aquesta empremta electrònica canvia, la qual cosa deixa clar que el fitxer ha estat modificat, i això en posa en dubte l'autenticitat.

Les tècniques forenses digitals també són fonamentals per a l'autenticació del contingut digital. Els experts utilitzen eines d'anàlisi d'imatges i vídeos per identificar artefactes o patrons visuals que no coincideixen amb el contingut original. Això pot incloure la detecció d'errors de compressió, anomalies en la il·luminació o les ombres, inconsistències en la resolució o en la qualitat de la imatge o problemes amb la sincronització del so i el moviment dels llavis en vídeos. Aquestes tècniques requereixen coneixements avançats i sovint s'utilitzen en casos en què l'evidència digital és clau, com en investigacions judicials o en la cobertura de notícies.

Un altre aspecte innovador és l'ús de cadena de blocs (*blockchain*) per a l'autenticació i la protecció del contingut digital. La tecnologia de cadena de blocs crea un registre distribuït i immutable que documenta, de manera transparent, qualsevol modificació d'un fitxer. Cada vegada que el contingut es modifica, s'actualitza el registre a la cadena de blocs, la qual cosa permet rastrejar-ne tots els canvis. Aquesta tecnologia és especialment útil en entorns en què la integritat de la informació és crucial, com la gestió de drets d'autor, la informació periodística o en processos legals.

Les tècniques com la cadena de blocs permeten rastrejar els canvis en el contingut digital i protegir-lo davant possibles manipulacions (Westerlund, 2019). La cadena de blocs no només garanteix l'autenticitat, sinó que també

actua com un sistema de protecció dels drets d'autor i les llicències digitals. Mitjançant contractes intel·ligents, es poden gestionar els drets d'ús i la distribució d'imatges o vídeos, cosa que ofereix un sistema de gestió transparent per als creadors i titulars dels drets.

Un enfocament prometedori per combatre els hipertrucatges és la combinació de diverses tècniques de detecció i autenticació, que aprofita els punts forts de cadascuna per crear sistemes més robustos i efectius. Aquesta tendència integra mètodes com el marcatge digital (marques d'aigua), l'anàlisi forense i els algorismes basats en aprenentatge automàtic (Megías *et al.*, 2022). La combinació d'aquestes tècniques pot millorar significativament la precisió i la fiabilitat dels sistemes de detecció.

Quan la tecnologia no és suficient per verificar l'autenticitat del contingut, es poden considerar testimonis i experts en peritatge digital. Aquests professionals poden presentar anàlisis tècniques en tribunals per verificar la integritat d'un fitxer digital. Els testimonis poden corroborar la procedència del contingut, la qual cosa, tot i no ser tan sòlida com una prova tècnica, pot ajudar a establir-ne l'autenticitat en un context legal.

En resum, la combinació de tècniques com la verificació de les metadades, les marques d'aigua, les signatures digitals, la cadena de blocs i les eines forenses digitals ofereixen una solució sòlida per autenticar continguts digitals i garantir que els vídeos o les imatges no hagin estat manipulats. En un món cada vegada més vulnerable a la desinformació i la manipulació digital, aquestes eines són essencials per mantenir la confiança en els continguts multimèdia que consumim diàriament.

Reptes tecnològics

El sorgiment de la tecnologia d'hipertrucatge ha estat objecte d'una anàlisi crítica, i se n'ha destacat el potencial disruptiu en sectors com els mitjans de comunicació i la política (Westerlund, 2019). Però els hipertrucatges plantegen una sèrie de reptes tecnològics i socials significatius que requereixen solucions sofisticades i multidisciplinàries.

A mesura que la tecnologia evoluciona, també ho fan els mètodes utilitzats per generar aquests continguts falsos, cosa que en complica la detecció i prevenció. Un dels principals reptes és la millora contínua del realisme. Els hipertrucatges s'han tornat extremament realistes gràcies a avenços en tècniques com les GAN i els autoencoders.

Les anomalies evidents en les primeres generacions d'hipertrucatges (desincronització labial, moviments corporals incoherents o errors en la textura de la pell) s'han reduït dràsticament. A més, les aplicacions modernes poden generar vídeos falsos en temps real, cosa que en fa gairebé impossible la verificació immediata.

Entre els desafiaments principals de les tècniques de detecció trobem, en primer lloc, el risc de falsos positius, en què contingut legítim es classifica incorrectament com a fals. Els falsos positius poden minar la credibilitat dels

mitjans digitals, i alimentar la desinformació i la desconfiança. A més, pot provocar que alguns continguts genuïns siguin eliminats o bloquejats de plataformes digitals.

En segon lloc, les eines de detecció d'hipertrucatges estan sovint orientades a experts o grans organitzacions. Això deixa els usuaris finals amb poques opcions per verificar contingut. De fet, moltes d'aquestes eines són complexes i poc accessibles per a la població general. Lligat amb aquest problema es dona el fet que quan els usuaris no entenen com funcionen les eines de detecció poden arribar a desconfiar dels resultats.

Finalment, la desconfiança generalitzada en el contingut digital pot tenir un impacte devastador en sectors com els mitjans de comunicació o la justícia, en què la credibilitat de les proves digitals és essencial (Colado, 2021).

Interacció dels usuaris amb les eines de detecció

Una dimensió essencial de l'èxit de les eines de detecció i autenticació d'hipertrucatges és la manera com els usuaris interactuen amb aquestes tecnologies. La confiança i la usabilitat són factors clau que determinen si aquestes eines s'adoptaran àmpliament i si tindran l'impacte desitjat en la lluita contra la desinformació i la manipulació digital. Malgrat els avenços tecnològics, moltes eines disponibles actualment són complexes i no sempre intuïtives per a l'usuari final, la qual cosa pot limitar-ne l'ús efectiu fora d'entorns altament especialitzats. Un dels principals reptes és com fer que aquestes eines siguin accessibles i comprensibles per a un públic ampli, des de professionals dels mitjans de comunicació fins a ciutadans preocupats per la integritat dels continguts digitals.

Segons Rosales *et al.* (2024), un sistema no només ha de ser precís, sinó també transparent i explicable. Això implica que els resultats oferts per les eines de detecció han de ser fàcilment comprensibles, amb informació clara sobre com s'ha arribat a una conclusió determinada, ja sigui per detectar un contingut manipulat o per validar-ne l'autenticitat. Aquesta explicabilitat fomenta la confiança en els sistemes i ajuda a evitar que es considerin «caixes negres» opaques.

A més, cal que els sistemes estiguin dissenyats amb un enfocament en la usabilitat. Això inclou interfícies intuïtives, temps de resposta ràpids i integracions amb plataformes populars, com les xarxes socials o els sistemes de gestió de contingut. Un exemple en podria ser la implementació d'extensions de navegador o aplicacions mòbils que analitzin contingut en temps real i informin l'usuari de manera immediata i clara si un vídeo, àudio o imatge pot ser un hipertrucatge. També és important que aquestes eines ofereixin diferents nivells de profunditat en els seus informes, des d'indicadors senzills per a usuaris no experts fins a detalls tècnics més avançats per a investigadors o professionals.

Un altre aspecte crucial és la confiança en els sistemes. La confiança en aquests sistemes es pot consolidar mitjançant

jaçant auditories independents que validin la precisió i l'eficàcia de les eines, així com mitjançant l'ús de conjunts de dades obertes que permetin a la comunitat científica i tècnica replicar els resultats i verificar la robustesa dels sistemes. A més, la integració d'elements de seguretat, com la privadesa dels usuaris i la protecció de dades personals, també és fonamental per garantir que aquestes tecnologies es percebin com a segures i responsables.

En resum, perquè les eines de detecció d'hipertrucatges siguin efectives i àmpliament utilitzades, han de ser no només tecnològicament avançades, sinó també dissenyades amb un enfocament centrat en l'usuari. La combinació de transparència, explicabilitat i usabilitat pot augmentar la confiança en aquestes tecnologies i promoure una adopció més àmplia, i ampliar-ne l'impacte positiu en la societat.

Recomanacions pràctiques per als lectors

Davant els riscos associats als hipertrucatges, és essencial dotar els lectors d'eines i estratègies pràctiques per identificar, gestionar i minimitzar l'impacte d'aquesta tecnologia en el seu dia a dia. A continuació, es detallen algunes recomanacions per protegir-se dels continguts manipulats i fomentar un consum responsable de la informació digital.

1. Verificació del contingut digital

Cal verificar l'autenticitat del contingut abans de compartir-lo o donar-lo per fiable. Per fer-ho:

- Assegureu-vos que el contingut prové d'una font fiable. Desconfieu de continguts virals sense una procedència clara.
- Corroboreu la informació amb altres mitjans o plataformes reconegudes.
- Utilitzeu eines de verificació com ara InVID, Forensically o FakeCatcher, que permeten analitzar vídeos i imatges per detectar manipulacions.

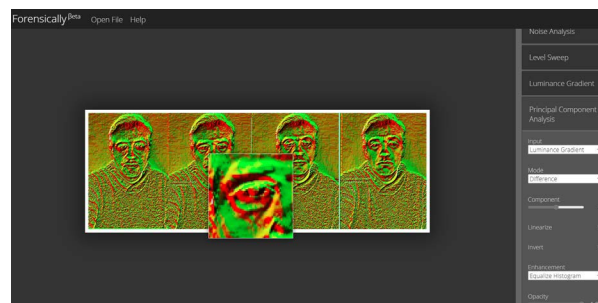


FIGURA 3. Anàlisi d'un hipertrucatge mitjançant l'aplicació Forensically. FONT: Elaboració pròpia.

Algunes xarxes socials i cercadors ofereixen funcionalitats per verificar la veracitat de continguts virals.

2. Identificació de signes de manipulació

Tot i que els hipertrucatges són cada vegada més sofisticats, hi ha indicadors que poden revelar possibles manipu-

lacions. Per exemple, es poden observar anomalies en la il·luminació o ombres que no concorden amb l'entorn. També és habitual que hi hagi errors subtils en les expressions facials o els moviments dels ulls.

Un altre aspecte complex a l'hora de fer un hipertrucatge són les textures facials, que acaben resultant molt artificials o excessivament uniformes.

Quant a l'àudio, es pot observar una petita desincronització entre el moviment dels llavis i el so i veus amb talls, distorsions o un to inusualment mecànic.

A més, tot i que no és sempre tan fàcilment observable, cal posar en dubte aquell contingut que genera reaccions emocionals extremes o apel·la a polèmiques sense proves sòlides.

3. Utilització d'eines tecnològiques

Els avenços en tecnologia també ofereixen eines per combatre els riscos dels hipertrucatges.

Per una banda, trobem programes que detecten anomalies en vídeos, com compressió irregular, patrons de píxels o inconsistències en la freqüència.

Com hem comentat, la tecnologia basada en cadena de blocs permet verificar l'autenticitat de continguts comparats en línia mitjançant registres immutables.

Per una altra banda, hi ha algunes extensions per als navegadors que permeten alertar sobre continguts sospitosos a les xarxes socials o a les plataformes de vídeo.

4. Educació digital i alfabetització mediàtica

Sens dubte, la millor manera de lluitar contra la falsedat és el coneixement. Una ciutadania formada és menys vulnerable a la desinformació i la manipulació. És altament recomanable:

- Participar en tallers, cursos o seminaris web sobre seguretat digital i detectar contingut fals.
- No creure's tot el que es veu en línia i qüestionar-se continguts sorprenents o polèmics.
- Que les escoles disposin d'eines i recursos per educar els estudiants sobre els riscos dels hipertrucatges i com identificar-los.

5. Contribucions a la lluita contra l'hipertrucatge

També podem contribuir a una regulació més robusta de la informació denunciant continguts sospitosos a les plataformes digitals perquè siguin revisats.

A més de mesures tècniques i pràctiques, és important que cada lector adopti una posició ètica respecte a la creació i difusió de continguts digitals i tingui cura de:

- Compartir només continguts verificats per evitar la propagació de desinformació.
- No utilitzar tecnologia per manipular continguts d'altres persones sense consentiment.
- Reflexionar sobre com el contingut digital influeix en les opinions i en els comportaments propis.

La millor defensa contra els hipertrucatges és una combinació de coneixement tecnològic, pensament crític i

consciència ètica. Amb una actitud proactiva i una educació digital sòlida, és possible protegir-se dels riscos de la manipulació digital i contribuir a una societat més segura i responsable.

Conclusió

Els hipertrucatges representen una revolució tecnològica que, tot i tenir un gran potencial creatiu i educatiu, planteja desafiaments crucials en l'àmbit de l'ètica, la privacitat i la seguretat. Les implicacions ètiques i socials dels hipertrucatges són profundes i multidimensionals, i afecten sectors com la política, l'entreteniment, els mitjans de comunicació o fins i tot les relacions personals.

Els hipertrucatges representen un repte per a la confiança social, ja que qüestionen la nostra capacitat per discernir entre la realitat i la falsedat en un món digital. Aquesta pèrdua de confiança té repercussions directes en la credibilitat de les institucions, especialment en el context de les notícies i els processos democràtics. Per exemple, els vídeos manipulats poden ser utilitzats per desacreditar figures polítiques o per manipular l'opinió pública, i erosionar així els fonaments de les democràcies modernes.

Una altra implicació significativa és la vulneració de la privacitat i la dignitat personal. Els hipertrucatges s'han utilitzat en nombrosos casos per crear contingut pornogràfic fals sense el consentiment de les víctimes, sovint dones, amb greus conseqüències psicològiques i reputacionals. Aquesta forma de violència digital no només afecta les víctimes directament implicades, sinó que també planteja preguntes més àmplies sobre com protegir els drets individuals en l'era digital. L'ús d'aquesta tecnologia per a la difamació i l'extorsió, així com la seva capacitat per semblar desconfiança entre comunitats, evidencia la seva potencialitat com a arma danyosa.

Les conseqüències socials dels hipertrucatges també inclouen l'augment de la desinformació, que pot ser explotada per actors malintencionats per crear divisions socials i culturals. La propagació de continguts manipulats genera confusió i dificulta la presa de decisions informades, tant en l'àmbit personal com col·lectiu. Aquesta desinformació massiva es veu amplificada per la velocitat i l'abast de les plataformes digitals, que sovint prioritzen el contingut viral sobre l'autenticitat. Això no només perjudica les víctimes directes dels hipertrucatges, sinó que també genera un entorn d'incertesa generalitzada que pot ser explotat per erosionar la confiança en els mitjans i en les institucions públiques.

Des d'una perspectiva ètica, els hipertrucatges plantegen qüestions sobre la responsabilitat dels creadors d'aquesta tecnologia i dels qui la fan servir. Si bé la innovació tecnològica pot tenir usos positius, com en el cinema o l'educació, el seu ús malintencionat posa de manifest la necessitat d'establir límits ètics i regulacions clares. Això inclou abordar preguntes sobre qui hauria de tenir accés a aquestes eines, com se'n pot prevenir l'ús indegut i quines sancions són apropiades per als infractors.

A més, la responsabilitat també recau en les plataformes digitals, que actuen com a vehicles de difusió dels hipertrucatges. Aquestes plataformes tenen l'obligació moral i social d'implementar mecanismes per detectar i eliminar contingut falsificat abans que esdevingui viral. Alhora, és crucial que aquestes iniciatives siguin transparents i respectin els drets dels usuaris, incloent-hi la privadesa i la llibertat d'expressió. La manca de compromís per part de les grans corporacions tecnològiques podria exacerbar els problemes associats als hipertrucatges, i convertir-los així en eines cada vegada més perilloses.

Finalment, les implicacions socials també inclouen l'impacte a llarg termini sobre la confiança en el contingut digital. A mesura que els hipertrucatges es tornen més sofisticats i difícils de detectar, existeix el risc que el públic desenvolupi una desconfiança generalitzada envers tot tipus de contingut digital, fins i tot el legítim. Això podria tenir conseqüències devastadores en sectors com la justícia, en què la credibilitat de les proves digitals és essencial, i en la comunicació, en què la fiabilitat de les notícies és fonamental per al funcionament d'una societat informada.

Abordar les implicacions ètiques i socials dels hipertrucatges requereix un esforç col·lectiu que inclogui la regulació legal, la responsabilitat tecnològica, l'educació digital i la sensibilització pública. Sense un enfocament integral, aquesta tecnologia pot desestabilitzar estructures socials fonamentals, amplificant desigualtats i posant en perill la confiança en el món digital. Només amb una resposta coordinada serà possible aprofitar els beneficis dels hipertrucatges alhora que se'n minimitzen els efectes perjudicials.

D'altra banda, la ràpida evolució de la tecnologia d'hipertrucatge exigeix una resposta contínua i coordinada. Les eines forenses digitals i les tècniques d'aprenentatge automàtic han de mantenir-se en desenvolupament constant per seguir el ritme de les noves tècniques de falsificació. En aquest sentit, la col·laboració entre experts en tecnologia, legisladors i investigadors és fonamental per anticipar i mitigar les noves amenaces que els hipertrucatges poden comportar en el futur.

No obstant això, la tecnologia per si sola no és suficient per contrarestar aquests desafiaments. També és necessari promoure una educació digital que permeti a la ciutadania identificar continguts manipulats i comprendre l'impacte potencial de les notícies falses o alterades. Al mateix temps, la creació d'infraestructures legals robustes

que penalitzin l'ús maliciós d'hipertrucatges i que protegeixin les víctimes és un pas indispensable per mantenir la integritat i la confiança en els mitjans digitals.

En resum, la tecnologia d'hipertrucatge obre un nou camp d'oportunitats, però també exigeix que la societat es doti d'eines tècniques, normatives i educatives per assegurar un ús ètic i responsable d'aquesta tecnologia potent.

Bibliografia

- COLADO, S. (2021). *Multiversos digitales: La tecnología como palanca evolutiva*. Barcelona: Universo de las Letras.
- GONG, D.; LI, Y.; WANG, F. (2020). «Forense de deepfakes: detección sintética per IA amb GANs convolucionals profundes». *Revista Internacional d'Intel·ligència Artificial*, 14 (3), p. 25-36.
- MEGÍAS, D.; KURIBAYASHI, M.; ROSALES, A.; CABAJ, K.; MAZURCZYK, W. (2022). «Architecture of a fake news detection system combining digital watermarking, signal processing, and machine learning». *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, 13 (1), p. 33-55.
- PARIS, B.; DONOVAN, J. (2019). «Deepfakes and cheap fakes: The manipulation of audio and visual evidence» [en línia]. Data & Society Research Institute. <<https://data.society.net/>> [Consulta: 25 abril 2024].
- ROSALLES, A.; MALANOWSKA, A.; KOOPAYEH ARAGHI, T.; KURIBAYASHI, M.; KOWALCZYK, M.; BLANCHE-TARRAGÓ, D.; MAZURCZYK, W.; MEGÍAS, D. (2024). «Trustworthiness and explainability of a watermarking and machine learning-based system for image modification detection to combat disinformation». A: *Proceedings of the 19th International Conference on Availability, Reliability and Security (ARES '24)* [en línia]. Association for Computing Machinery, p. 1-10. <<https://doi.org/10.1145/3664476.3670934>>.
- SHAD, H. S.; LÓPEZ, J. A.; KUMAR, R. (2021). «Anàlisi comparativa del mètode de detecció d'imatges deepfake utilitzant xarxes neuronals convolucionals». *Intel·ligència Computacional i Neurociències*, 18 (4), p. 102-118.
- TERMCAT (2024). «Hipertrucatge» [en línia]. <<https://www.termcat.cat/ca/diccionaris-en-linia>>.
- WESTERLUND, M. (2019). «El sorgiment de la tecnologia hipertrucatge: una revisió». *Technology Innovation Management Review*, 9 (11), p. 39-52.