EINES LEGALS I TECNOLÒGIQUES PER COMBATRE LES FAKE NEWS AL CIBERESPAI

Màster Universitari de Ciberdelinqüència

Autor: Eduard Emilià Vilà Méndez

Director: Andreu Van Den Eynde Adroer

Semestre: Febrer – Juliol 2023

15 de juny de 2023

Agraïments

Vull agrair a la Maite, la meva dona, per acompanyar-me i recolzar-me de forma incondicional en tots els projectes que he anat emprenent, incloent aquest Màster i Treball Fi de Màster. Només ella sap el temps i l'esforç que he hagut d'invertir. Sense el seu suport, paciència i ajuda no ho podria haver aconseguit.

Al meu petit Aleix, que des de fa poc més d'un any que il·lumina tot el que hi ha al seu voltant. És una joia veure'l créixer dia a dia.

A la meva família per haver estat allí sempre, en les bones i en les no tan bones, i ser aquell suport que tota persona necessita per tirar endavant en la vida.

Al meu director, l'Andreu, per ajudar-me a triar el tema del treball, acceptar-ne ser el director, accedir sempre de bon grat les propostes que li he fet, i donar-me tot de consells i observacions que m'han permès acabar d'enfocar i polir la versió final del treball.

També vull agrair especialment a l'Il·lustríssim Miguel Ángel Aguilar García, recentment nomenat Fiscal de Sala de Delictes d'Odi i Discriminació, per haver respost tan ràpidament als correus electrònics que li vaig enviar però, sobretot, pel tracte proper i cordial que en tot moment em va dispensar.

Finalment, però no menys important, no puc ni vull deixar passar l'ocasió de retre un humil i sentit record a la meva mare, la Leticia. Una cabuda, afectuosa i orgullosa puertorriqueña a la que una cruel malaltia li va privar de la seva memòria i una implacable pandèmia ens va arrabassar del seu costat.

ÍNDEX

1. INTRODUCCIÓ

Què són les *fake news*? Podem trobar innumerables definicions que cerquen explicar-ne el seu significat. Si volem mostrar una de senzilla però complerta, disposem de la que ens proporciona el professor en filosofia Axel Gelefert, que descriu *fake news* com "*la presentació deliberada de (normalment) afirmacions falses o enganyoses com a notícies, en les quals les afirmacions són dissenyades per resultar enganyoses*" (Gelfert, 2018). Per la seva banda, la Fiscalia General de l'Estat les equipara amb l'expressió "*desinformar*" que, segons la Reial Acadèmia Espanyola, consisteix en "*donar informació intencionadament manipulada al servei de certs fins*" o "*donar informació insuficient o ometre-la*", mentre que per la Comissió Europea *fake news* són "*aquelles que inclouen informació no veraç o manipulada. Poden tenir aparença de notícies clàssiques, però el contingut persegueix enganyar el públic*"[1].

En aquest sentit, l'any 2017 "*fake news*" va ser escollida paraula de l'any pel diccionari britànic Collins[2], que la defineix com aquella "*informació falsa, sovint sensacionalista, difosa sota l'aparença de la notícia*". Collins va basar la seva elecció per la "*presència omnipresent*" que *fake news* va tenir en el transcurs d'aquell any, en el qual els seus lexicògrafs, que supervisen una base de dades formada per més de 4.500 milions de paraules, van constatar com l'ús d'aquest terme havia augmentat fins un 365% respecte l'any anterior. Buscant els motius d'aquesta explosió sobtada, es pot afirmar que aquesta està estretament lligada amb els esdeveniments relacionats amb el referèndum pel Brexit al Regne Unit i amb les eleccions nord-americanes ganyades per Donald Trump, tots dos fets ocorreguts l'any 2016 (Parra Valero & Oliveira, 2018).

Si cerquem fer una classificació de les *fake news* segons la seva intencionalitat, autors com Claire Wardle i Hossein Derakhshan les divideixen en tres categories diferenciades: la *informació errònia* o *misinformation*, que és aquella informació que, tot i resultar falsa, la persona que la difon molts cops creu que és certa i no ha estat creada amb la intenció de fer mal; la *desinformació* o *disinformation*, tota informació que resulta ser falsa i que, a més, la persona que la difon coneix de la seva falsedat i té la intenció de causar un mal (en aquest cas estem parlant d'una mentida creada de forma deliberada i

[1] https://spain.representation.ec.europa.eu/noticias-eventos/noticias-0/como-combatir-las-fake-news-2022-02-28_es

[2] https://blog.collinsdictionary.com/language-lovers/collins-2017-word-of-the-year-shortlist/

intencionada dirigida a un públic concret); i, finalment, la *informació incorrecta* o *mal-information*, que es refereix a la informació que es basa en la realitat, però s'empra deliberadament per fer mal a una persona, col·lectiu, organització o país, però també representa tota aquella informació genuïna que es comparteix per fer mal, en molts casos traslladant informació privada a l'esfera pública (Wardle & Derakhshan, 2018).

Cercant sobre els seus orígens, resulta obvi que les *fake news* no són, ni molt menys, un fenomen nou, ja que el transcurs de la història de la humanitat s'ha fet ús en múltiples ocasions de la mentida i la informació deliberadament tergiversada amb la intenció d'assolir objectius i finalitats concrets relacionats, normalment, amb l'assoliment de poder. Un dels primers exponents és "L'art de la guerra" de Sun Tzu, redactat entre els segles IV i VI a.C (Sun-Tzu, 2016), el qual defensa que l'ideal és guanyar sense haver de lluitar, ja que la guerra es basa en l'engany i la confusió de l'enemic. Sun Tzu afirma que "*l'art de la guerra es basa en l'engany"* i afegeix que "*el principal engany que es valora en les operacions militars no es dirigeix només als enemics, sinó que comença per les pròpies tropes, per fer que el segueixin a un sense saber on van*". Un altre exemple d'aquest tipus de manipulació informativa històrica el trobem en els "*Cronicons*" espanyols del segle XVI, els quals, sota l'aparença de cròniques fidedignes informatives o històriques, resultaven ser documents deliberadament redactats i arranjats amb tot tipus de suposicions, imaginacions i fabulacions (Fernández Fernández, 2020). Finalment, un cas més proper en el temps el trobem amb la que es va anomenar "*La gran mentida de la lluna*" de l'any 1835, on el diari nord-americà "The Sun" va publicar durant sis dies com un científic britànic havia albirat vida intel·ligent a la lluna gràcies al seu potent telescopi. Aquesta informació falsa i sensacionalista sota la disfressa d'una notícia veritable va causar un gran impacte als Estats Units i a Europa gràcies a la concatenació de tres factors: l'aparició de les premses d'alta capacitat, la caiguda del preu dels diaris i l'arribada dels nous i més ràpids mitjans de transport com els trens i els vaixells (Salas Abad, 2019).

Centrant-nos en el present, els canvis socials viscuts les darreres dècades arrel de l'aparició de les tecnologies d'informació en general i Internet en particular tenen el seu reflex en la criminalitat com a fenomen social que és, cosa que ha portat a l'aparició d'un nou tipus de delinqüència relacionada amb el ciberespai (Llinares, 2012). Aquests canvis, que han desembocat en un ús massiu i universal d'Internet i les TIC en general, i de les grans plataformes, xarxes socials i aplicacions de missatgeria instantània en particular,

han facilitat enormement la proliferació de tot tipus de ciberdelinqüència, entre la que s'inclou aquella que basada en el robatori, l'ús, l'explotació i la manipulació de la informació. És en aquest context on les *fake news* han trobat en el ciberespai un nou espai en el qual explotar noves finalitats il·lícites o delictives i que, a més, ha esdevingut un mitjà idoni on divulgar-se i propagar-se de forma ràpida i eficaç, podent arribar a un públic potencial format per milions de persones. Una revolució equiparable a la que es va viure en el seu moment l'any 1835.

En aquest context, l'explosió tecnològica viscuda arrel de la popularització en l'ús d'Internet i les TIC ha propiciat que les *fake news* hagin derivat en campanyes massives de desinformació, emprades com a eina política per manipular l'opinió pública, erosionar l'estabilitat dels Estats i les seves institucions, debilitar la confiança de la ciutadania en aquests, i desacreditar els pensaments i/o plantejaments polítics oposats (Morejón-Llamas, Martín-Ramallal, & Micaletto-Belda, 2022). Exemples els trobem, alguns ja esmentats anteriorment, en les campanyes de desinformació relacionades amb les eleccions presidencials dels Estats Units d'Amèrica del 2016 guanyades per Donald Trump i del Brasil el 2018 guanyades per Jair Bolsonaro, així com el referèndum del Brèxit del 2016 (Rodríguez Pérez, 2019).

D'altra banda, en la generació i difusió de *fake news* relacionades amb la pandèmia de la COVID-19, les grans plataformes d'Internet i les xarxes socials van esdevenir un escenari propici per on divulgar fabricacions informatives, manipulacions gràfiques, teories conspiratòries, continguts intencionadament descontextualitzats i falsedats de tot tipus (Salaverría, et al., 2020). Això va desembocar en els anomenats moviments negacionistes i antivacunes que, emprant el gran altaveu que els oferia Internet i les TIC, pretenien evidenciar un suposat objectiu ocult així com efectes no desitjats de les vacunes contra la COVID-19, i també exposaven suposats beneficis de tractaments alternatius presumptament miraculosos. Tot això amb el perill per la salut pública que aquests continguts podien arribar a comportar.

Però no només les *fake news* s'han emprat en el context de grans campanyes de desinformació, també han estat, per nombrar un exemple clar, l'eina per vessar discursos d'odi envers persones, grups i col·lectius concrets, com és el cas del discurs islamòfob divulgat a les xarxes socials en el que es fomenta la narrativa antiimmigratòria i

s'estigmatitza tot el relacionat amb la cultura i religió musulmana (Fuentes-Lara & Arcila-Calderón, 2023).

Per tot això, resulta evident que el gran perill de les *fake news* radica en el fet del poder que poden tenir per modificar la opinió pública i afectar l'agenda mediàtica tradicional. Tot això agreujat pel fet que han demostrat tenir un gran poder de propagació, cosa que les fa, fins i tot, tenir més impacte que la informació verdadera (Fernández-García, 2017). A més, les *fake news* poden ser emprades, com es veurà en aquest treball, per cometre una diversitat de delictes i per amenaçar la seguretat dels ciutadans, la salut pública, els béns públics i, fins i tot, els processos electorals i democràtics.

1.1. Qüestió i àmbit de recerca

Davant d'aquest fenomen, cal fer-se la següent pregunta: és possible lluitar de forma eficaç contra les *fake news* divulgades a través del ciberespai? El present treball busca respondre a aquesta pregunta enfocant-ho des de una doble vessant: la lluita legal i la lluita tecnològica contra les *fake news*.

Amb relació a la lluita legal, la recerca es centra en les eines que les normatives espanyola i europea ofereixen per lluitar contra les *fake news*. Es pretén, doncs, cercar i llistar aquelles conductes consistents en la creació i divulgació de *fake news* que poden resultar delictives i, consegüentment, penades segons les legislacions espanyola i europea. Per tant, l'estudi té com objectiu descriure i mostrar l'encaix legal que es pot trobar en relació a la creació i divulgació de *fake news*, especialment aquelles que hagin estat elaborades i propagades emprant com a mitjà Internet i les TIC.

A nivell espanyol, l'estudi es centra en els delictes contemplats en el **Codi Penal** en els quals es pot incórrer quan s'elaboren i/o difonen *fake news*. Queda, per tant, fora de l'àmbit d'aquest treball l'estudi les possibles responsabilitats civils, com seria, per exemple, el cas en que es produís la vulneració i atemptat contra el dret a l'honor, intimitat personal i familiar i a la pròpia imatge regulada a l'article **18.1** de la **Constitució Espanyola**[3] (en endavant CE) i a la **Llei Orgànica 1/1982**[4], de 5 de maig.

[3] https://www.boe.es/buscar/act.php?id=BOE-A-1978-31229
[4] https://www.boe.es/buscar/act.php?id=BOE-A-1982-11196

A nivell Europeu, l'estudi vol mostrar com la lluita contra les *fake news* ha esdevingut un dels desafiaments més cabdals als que les institucions europees s'han hagut d'afrontar els darrers anys, sobretot amb les conseqüències derivades arrel de l'impacte causat per la seva propagació a través del ciberespai. En aquest sentit, l'estudi centra el seu interès en el recentment aprovat **Reglament (UE) 2022/2065** o **Llei de Serveis Digitals**[5] i en el **Codi de bones pràctiques de la Unió en matèria de desinformació**[6] que, de forma conjunta, han d'esdevenir la punta de llança per lluitar-hi de forma eficaç.

Pel que fa a la lluita tecnològica, la recerca es centra, sobretot, en aquelles eines i recursos que es troben disponibles per a qualsevol internauta, independentment de quin sigui el nivell dels seus coneixements informàtics. Per això, s'estudien aquells recursos que Internet i les noves tecnologies ofereixen en forma d'aplicacions gratuïtes i pàgines web, posant especial atenció als recursos **OSINT**, un conjunt de tècniques i eines emprades per recopilar informació i analitzar-la per obtenir coneixement i que, entre altres utilitats, poden ajudar a detectar i combatre aquells continguts que poden ser catalogats com a *fake news*. Una base que resulta d'interès la formen totes aquelles eines, ajudes i guies que des dels mateixos organismes i autoritats oficials, estatals i europeus, faciliten a qualsevol persona que hi estigui interessada.

En aquest punt, resulta d'especial interès la figura dels **verificadors de fets** o **fact-checkers**, una branca del periodisme que utilitza eines tecnològiques i periodístiques per tal de desemmascarar notícies inexactes, errònies, ambigües o, directament, falses. Per aquest motiu, esdevenen un actor clau que juga un paper molt rellevant en la lluita contra les *fake news*. És per això que el 5 de desembre de 2018, a l'informe anomenat **La lluita contra la desinformació en línia: Un enfocament europeu**[7], la Comissió Europea es va comprometre a crear una xarxa independent de verificadors de fets per lluitar contra les *fake news* divulgades pel ciberespai. A més, la normativa europea recentment aprovada ha inclòs aquesta figura en el seu articulat.

[5] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-services-act-ensuring-safe-and-accountable-online-environment_es
[6] https://digital-strategy.ec.europa.eu/es/policies/code-practice-disinformation
[7] https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2018:0794:FIN:ES:PDF

Aquests verificadors de fets s'han erigit com una figura cabdal per lluitar contra les *fake news*, funcionant com un autèntic antídot a la desinformació que viatja per les considerades "artèries digitals" d'Europa, en tant que resulten de vital importància per desemmascarar les *fake news* i proveir als ciutadans d'aquelles proves fefaents que els ajudin a distingir entre la informació que és veritat i la que està manipulada o directament resulta ser mentida (Valcárcel Siso, Carrascal Domínguez, Pintado, & Nicolás, 2020).

1.2. Metodologia emprada

Entre les diverses alternatives existents, per dur a terme el present treball s'ha optat per emprar una metodologia basada, per una banda, en una **investigació empírica qualitativa**, que consisteix en la recopilació de dades amb l'objectiu de descriure les qualitats d'un fenomen, obtenint així un coneixement i enteniment profund i precís d'aquest (Palacios, 2006) i, per l'altra, en una **investigació formal-jurídica o dogmàtica** (Odar, 2016), que consisteix en estudiar les estructures del dret objectiu, és a dir; la norma jurídica i l'ordenament jurídic.

Així doncs, per mostrar la lluita legal contra les *fake news*, s'ha optat per realitzar una **investigació formal-jurídica o dogmàtica**. Concretament, es mostra quin és l'encaix legal que s'ha trobat a nivell espanyol i europeu que permet poder lluitar eficaçment contra les *fake news*. A més, per complementar l'estudi, a nivell europeu es realitza també una **recerca documental** entre els diversos informes, codis i plans que el Parlament Europeu i la Comissió Europea han aprovat en relació aquest tema.

D'altra banda, per descriure el fenomen de les *fake news* així com la lluita tecnològica per combatre-les, s'ha optat per realitzar una **investigació empírica qualitativa** per mitjà d'una recerca i estudi **documental bibliogràfics** (Gómez, 2011). L'objectiu ha estat tenir un bon coneixement sobre les investigacions i publicacions relacionades amb aquests temes per tal de seleccionar, organitzar i analitzar les més adients, i així poder descriure aquest fenomen de la forma més acurada i veraç possible. Per tal de complementar la investigació, també s'han consultat aquells documents i informació publicats en portals web d'entitats i administracions públiques i oficials, tant estatals com europees.

1.3. Fonts emprades per la recollida d'informació

En la recerca per obtenir les dades que han fomentat la **recerca documental** duta a terme, s'han emprat webs i portals d'organismes i administracions públiques, així com portals d'investigació i producció científics com *Dialnet*, *Recolecta* i, sobretot, *Google Scholar*. També ha resultat de gran utilitat la biblioteca de la UOC.

Per verificar la validesa i utilitat entre les fonts d'informació disponibles, s'ha realitzat una avaluació prèvia basant-se en la seva **autenticitat**, **credibilitat**, **representativitat** i **significat** (Scott, 2014). A més, resulta cabdal que les fonts de dades siguin objectives i no parteixin d'una visió condicionada o partidista. En relació al seu origen, les fonts que s'han emprat per recopilar tota la informació han estat, bàsicament, **fonts primàries** provinents de normatives legals i d'estudis, informes i articles d'investigació acadèmics. També s'ha fet ús de **fonts secundàries** provinents, fonamentalment, de diccionaris, articles de revistes i treballs en els quals s'han analitzat i/o interpretat investigacions anteriors. Finalment, també s'han consultat **fonts oficials primàries** disponibles als portals web d'organismes i administracions públiques, tant a nivell estatal com a nivell europeu.

2. LLUITA LEGAL CONTRA LES *FAKE NEWS*

La vessant legal esdevé un pilar fonamental en la batalla contra les *fake news*. Tot i això, resulta important destacar que la creació i difusió de *fake news*, per si mateixos, no constitueixen cap tipus de conducta delictiva. Dependrà del contingut que es creï i/o difongui a través d'una *fake news* i, sobretot, la intenció amb la que es publiqui i/o divulgui, així com el resultat potencialment o efectivament produït, per tal que es produeixi l'encaix en les normes legals existents que, en funció de la infracció que s'hagi donat, poden comportar penes de multa i, fins i tot, de presó.

Per això, es volen mostrar aquelles eines legals que, a nivell estatal i europeu, s'empren per lluitar contra la creació i difusió de *fake news*, atenent a les repercussions i conseqüències que poden arribar a generar, i centrant l'interès en aquelles que utilitzen el ciberespai com a mitjà de divulgació.

2.1. Lluita legal contra les *fake news* a Espanya

En relació al tractament que la normativa espanyola, concretament el seu **Codi Penal**, ha realitzat per tal de combatre la propagació de *fake news*, cal remarcar que la COVID-19 va esdevenir un punt d'inflexió per la gran quantitat i diversitat de *fake news* relacionades amb la pandèmia que van arribar a propagar-se pel ciberespai i, més concretament, a través de les grans plataformes, xarxes socials i aplicacions de missatgeria instantània.

Per aquests motius, la Fiscalia General de l'Estat (en endavant FGE) va publicar el 20 d'abril de l'any 2020 el **Repertorio de actuaciones FGE**[8], per tal de donar resposta des de diversos fronts als problemes derivats de la pandèmia de la COVID-19. Entre els diversos estudis, informes i anàlisis realitzats per la Secretaria Tècnica de la FGE, resulta d'interès per aquest treball l'estudi sobre les repercussions i implicacions penals de la difusió de *fake news*. En aquest informe, titulat **Tratamiento penal de las Fake News**[9], la Secretaria Tècnica de la FGE identifica una sèrie de delictes tipificats al **Codi Penal** espanyol en els quals es pot incórrer en el cas que s'elaborin i/o divulguin *fake news* de manera intencionada. En aquest sentit, l'informe classifica i relaciona els delictes amb les *fake news* depenent de a qui vagin referides i amb la intencionalitat que siguin divulgades, resultant els que seguidament es tracten.

2.1.1. Delictes d'Odi

Abans de res, cal destacar que, segons les definició de delictes d'odi realitzada per la Organització per la Seguretat i la Cooperació a Europa (**OSCE**) [10] i la facilitada per la Comissió Europea contra el Racisme i la Intolerància (**ECRI**) sobre el discurs d'odi[11], al **Codi Penal** espanyol els delictes d'odi pròpiament dits es centren, fonamentalment, en la circumstància agreujant de l'article **22.4 CP**, mentre que els delictes als quals es refereix la Secretaria Tècnica de la FGE es centren en la figura delictiva regulada a l'article **510 CP**, que és la relativa als delictes de provocació o incitació a l'odi i la discriminació. En aquest sentit, la **STS 1070/2019**, de 2 d'abril estableix que "*el bé jurídic protegit pel tipus*

[8] https://www.fiscal.es/documents/20142/402453/Recopilatorio+de+actuaciones+de+la+FGE+de+20+de+abril.pdf/b3b4460d-6295-307a-1cfc-3626a77da125?version=1.0
[9] https://www.icab.es/export/sites/icab/.galleries/documents-noticies/tratamiento-penal-de-las-fake-news-fiscalia-general-del-estado.pdf
[10] https://www.osce.org/files/f/documents/6/b/502275.pdf
[11] https://rm.coe.int/ecri-general-policy-recommendation-n-15-on-combating-hate-speech-adopt/16808b7904

penal de l'article 510 és la dignitat de les persones, i el col·lectiu de persones, als que per la seva especial vulnerabilitat el Codi atorga una protecció específica en el mencionat article"[12].

Així doncs, sense entrar en el debat sobre la frontera que hi pot haver entre el que pot suposar un delicte de provocació a l'odi i la discriminació regulat a l'article **510 CP** i la llibertat d'expressió de l'article **20.1.a CE**, els tipus delictius als que es refereix l'article **510.1 CP**, castigats amb penes de presó de un a quatre anys i multa de sis a dotze mesos, són aquells referits a les conductes públiques que fomentin, promoguin, incitin, de forma directa o indirecta a l'odi, hostilitat, discriminació o violència (article **510.1.a**), qui produeixi, elabori o posseeixi amb finalitat de distribuir, qui faciliti a terceres persones l'accés, distribueixin, difonguin o venguin qualsevol tipus de material que resulti idoni per fomentar, directa o indirectament, l'odi, hostilitat, discriminació o violència (article **510.1.b**), i la negació, trivialització o enaltiment del genocidi, de lesa humanitat o contra les persones i altres conductes similars o s'enalteixi els seus perpetradors quan així es promogui o afavoreixi un clima de violència, hostilitat, odi o discriminació (article **510.1.c**), quan qualsevol d'aquestes actituds vagin dirigides contra qualsevol dels grups mencionats.

D'altra banda, l'article **510.2 CP** castiga amb penes de sis mesos a dos any i multa de sis a dotze mesos aquelles conductes consistents en menyscabar i lesionar la dignitat de les persones mitjançant accions que comportin humiliació, menyspreu o descrèdit o produeixin, elaborin, posseeixin amb la finalitat de distribuir, facilitin a terceres persones accés, distribueixin, difonguin o venguin qualsevol tipus de material o suport que pel seu contingut resultin idonis per lesionar la dignitat de les persones per representar una greu humiliació, menyspreu o descrèdit de qualsevol dels grups mencionats (article **510.2.a**), així com aquelles que constitueixin un enaltiment o justificació dels delictes que haguessin estat comesos contra un grup, part d'aquest o persona per les raons llistades, o a qualsevol que hagi participat en la seva execució (article **510.2.b**). A més, l'article **510.2 CP** contempla una modalitat agreujada que castiga amb una pena d'un a quatre anys de presó i multa de sis a dotze mesos quan amb les actituds descrites es promogui o afavoreixi un clima de violència, hostilitat odi o discriminació envers els mencionats

[12] https://www.poderjudicial.es/search/AN/openDocument/3ed8ea8562fa2619/20190409

grups. De manera similar, l'article **510.4 CP** estableix que quan els fets resultin idonis per alterar la pau pública o crear un greu sentiment d'inseguretat o temor entre els integrants del grup, s'imposarà la pena en la seva meitat superior, que podrà ser elevada fins la superior en grau. Finalment, l'article **510.5 CP** tracta sobre les penes d'inhabilitació especial per a professió o oficis educatius que venen associades a la pena principal.

Respecte a les conseqüències penals d'haver comés aquests delictes per mitjà d'Internet i les TIC, l'article **510.3 CP** estableix que les penes dels dos apartats anteriors s'imposaran en la seva meitat superior quan els fets s'haguessin dut a terme a través d'un mitjà de comunicació social, Internet o les TIC, de tal manera que es fes accessible a un elevat nombre de persones. Per la seva banda, l'article **510.6 CP** especifica que quan el delicte s'hagués comès per mitjà de les TIC, s'acordarà la retirada dels continguts i, en els casos que, per mitjà d'un portal d'accés a Internet o servei de la societat de la informació, es difonguin exclusiva o de forma destacada continguts consistents en discurs d'odi, s'ordenarà el bloqueig de l'accés o la interrupció del servei.

2.1.2. Delictes de descobriment i revelació de secrets

El segon tipus penal que, segons l'informe de la Secretaria Tècnica de la FGE, pot integrar la difusió de *fake news*, es centra en el cas que la difusió vagi acompanyada de la revelació de dades personals autèntiques. Davant d'aquest fet, especifica que pot existir un concurs amb el delicte de descobriment i revelació de secrets de l'article **197.3 CP** així com els agreujants dels articles **197.5** y **197.6 CP**. Cal remarcar en aquest punt que, com en l'informe no queda directament especificat, s'ha d'entendre que, en no existir pròpiament el delicte de difusió de *fake news*, es refereix a que es produeixi un concurs de delictes entre la resta de delictes enumerats en l'informe. Per exemple, un delicte que constituís un delicte de provocació o incitació a l'odi en el qual es publiquessin dades personals verídiques de la víctima.

Amb això, l'article **197.3 CP** es refereix a la difusió no autoritzada de secrets o dades personals, esdevenint, per una banda, un tipus agreujat dels articles **197.1** i **197.2 CP** que castiga amb penes de presó de dos a cinc anys si, a banda de les conductes tipificades en aquests articles, es produeix la difusió, revelació o cessió a tercers de qualsevol arxiu o document personal i privat. D'altra banda, castiga amb penes de presó

d'un a tres anys i multa de dotze a vint-i-quatre mesos aquell qui, amb coneixement del seu origen il·lícit i sense haver pres part en el seu descobriment, realitzi la conducta descrita al anteriorment. En aquesta ocasió, doncs, es castiga una conducta independent on l'autor no participa en cap dels delictes descrits en els articles **197.1** o **197.2 CP**, sinó que, coneixent l'origen il·lícit de les dades, fets descoberts o les imatges captades, els difon, revela o cedeix a tercers.

L'article **197.5 CP** és un tipus agreujat on s'especifica que s'imposaran les penes previstes en la seva meitat superior quan els fets constitutius d'espionatge (d'acord amb els apartats anteriors) afectin dades de les denominades "sensibles" o la víctima fos menor d'edat o persona amb discapacitat necessitada d'especial protecció. Aquest tipus de dades son les que fan referència a la ideologia, religió, salut, origen racial o vida sexual, la revelació de les quals porta implícit el perjudici previst pel tipus penal. Per la seva banda, l'article **197.6 CP** destaca que si els fets es realitzen amb fins lucratius, s'imposaran les penes previstes en la seva meitat superior. Si a més afecten a les dades del l'article **197.5 CP**, la pena serà de presó de quatre a set anys.

En aquest punt, cal esmentar la crítica que alguns juristes han realitzat respecte l'equiparació que l'article **197.3 CP** fa sobre la cessió a tercers de dades relatives a la intimitat de les persones respecte la seva difusió indiscriminada i incontrolada a través d'una *fake news*, sobretot si és pel ciberespai. Per això, hi ha juristes que defensen la necessitat de proposar un nou tipus agreujat que reculli, seguint el principi de proporcionalitat, la major gravetat d'aquestes conductes (Mora Díez, 2020).

2.1.3. Delictes contra la integritat moral

En aquest cas, l'informe de la Secretaria Tècnica de la FGE estableix que en casos d'extrema gravetat i essent la víctima una persona individual, les *fake news* podrien integrar-se en el delicte contra la integritat moral de l'article **173.1 CP**, que castiga amb penes de sis mesos a dos anys de presó aquell que infringeixi a una altra persona un tracte degradant, menyscabant greument la seva integritat moral. Per tant, deixa clar que el tracte degradant ha de menyscabar greument la integritat moral, cosa que exclouria els supòsits de menor entitat que, molt probablement, quedarien integrats en els delictes d'injúries i calúmnies. En aquest sentit, la **STS 3270/2022**, de 15 de setembre estableix que els elements que configuren el concepte d'atemptat contra la integritat moral són "*a)*

Un acte de clar i inequívoc contingut vexatori per al subjecte passiu, b) La concurrència d'un patiment físic o psíquic, i c) Que el comportament sigui degradant o humiliant amb especial incidència en el concepte de dignitat de la persona-víctima"[13].

Per la seva banda, els paràgrafs segon i tercer d'aquest article castiguen amb la mateixa pena aquells qui, en l'àmbit de qualsevol relació laboral o funcional i prevalent-se de la seva relació de superioritat, realitzin contra un altre de forma reiterada actes hostils o humiliants que, sense que arribin a constituir un tracte degradant, suposin un greu assetjament contra la víctima, o tinguin com objectiu impedir el legítim gaudi de l'habitatge. Tanmateix, tot i que, tal i com queda reflectit en l'informe de la FGE, és possible la comissió d'aquests delictes mitjançant la creació i divulgació de *fake news*, tant les conductes d'assetjament laboral com immobiliari emprant l'ús de *fake news* resulten de difícil comissió ja que difícilment s'arribarà a produir aquesta situació a la pràctica.

Finalment, tot i que a l'informe de la FGE no s'esmenta, es creu adient tractar el delicte de l'article **172 ter 5 CP**, que castiga amb penes de presó de tres mesos a un any o multa de sis a dotze mesos el que, sense consentiment del seu titular, utilitzi la imatge de la persona per fer anuncis o obrir perfils falsos en xarxes socials, pàgines de contacte o qualsevol mitjà de difusió pública ocasionant-li una situació d'assetjament, fustigació o humiliació. En aquest cas si que es troba una clara connexió entre la conducta descrita i la difusió de *fake news* a través del ciberespai en tant que emprar la imatge de la persona o crear-ne un perfil fals molt probablement aniran acompanyats de la divulgació d'informació falsa sobre aquesta persona.

2.1.4. Delictes de desordres públics

En aquesta ocasió, l'informe de la Secretaria Tècnica de la FGE estableix que quan les *fake news* continguin missatges d'alarma, atemptats terroristes o catàstrofes que impliquin situacions de perill per la societat o facin necessari l'auxili i l'activació dels serveis d'emergència, l'afirmació falsa o simulació podria ser constitutiva del delicte de desordres públics de l'article **561 CP**, que castiga amb penes de tres mesos i un dia a un any de presó i multa de tres a divuit mesos qui afirmi situacions de perill per la comunitat

[13] https://www.poderjudicial.es/search/AN/openDocument/611ec8fb42f82145a0a8778d75e36f0d/20220923

o perill de sinistre a partir de la qual sigui necessari prestar auxili a un altre i amb això provoqui la mobilització dels serveis d'emergència, i/o l'article **562 CP**, que preveu penes d'inhabilitació en el cas qui cometés els fets descrits estigués constituït en autoritat.

Tanmateix, el que es descriu a l'informe de la Secretaria Tècnica de la FGE no es correspon amb el redactat de l'article **561 CP**, ja que, segons aquest darrer, per tal que existeixi la conducta típica es necessari que es produeixi la mobilització dels serveis d'emergència, cosa que en l'informe no resulta imprescindible, en tant que n'hi hauria prou que la *fake news* simulés, falsament, una situació de perill per la societat.

Per tant, al parer del que redacta el present treball i contràriament del que afirma l'informe de la Secretaria Tècnica de la FGE, per tal que es produeixi l'encaix penal d'aquest delicte, és necessari que la *fake news* tingui una autèntica aparença veraç quan, en realitat, és totalment fal·laç. A més, aquell qui la difongués ho hauria de fer amb total coneixement de la seva falsedat assumint alhora que, com a mínim, existís una possibilitat real de l'activació i mobilització efectiva dels serveis d'emergència. Amb això, centrant-se en el supòsit que s'estudia en el present treball, la difusió de *fake news* a través del ciberespai, cal destacar la dificultat de la comissió efectiva d'aquest tipus de delicte, ja que resulta altament improbable que una *fake news* divulgada, per exemple, per una xarxa social, pogués arribar a tenir un impacte tal que causés la mobilització dels serveis d'emergència, en tant que seria relativament fàcil contrastar-la i verificar-ne la seva autenticitat abans que aquests fossin efectivament activats. Un exemple pràctic el trobem amb els hoax que, tal i com els descriu l'Institut Nacional de Ciberseguretat o **INCIBE** "*són fake news creades sobre temes d'interès o d'actualitat per generar una alarma social o atreure l'atenció del nombre més gran d'usuaris possibles, per tal de generar desinformació o bé reconduir cap a algun altre tipus de frau*"[14]. En aquest sentit, tot i que els hoax tenen l'objectiu de generar alarma social, molt difícilment implicaran l'activació i mobilització dels serveis d'emergència, cosa que, a la pràctica, impossibilita que es pugui aplicar el delicte de desordres públics de l'article **561 CP**.

[14] https://www.incibe.es/aprendeciberseguridad/hoax-bulo#:~:text=El%20Concepto,alg%C3%BAn%20otro%20tipo%20de%20fraude

2.1.5. Delictes d'injúries i calúmnies

Continuant amb l'informe de la Secretaria Tècnica de la FGE, aquest especifica que les *fake news* poden integrar els delictes d'injúries i calúmnies dels articles **209** i **206 CP** respectivament. Així doncs, el delicte d'injúries, recollit a l'article **208 CP**, estableix que es considera injuria tota aquella acció o expressió que lesioni la dignitat d'una altra persona, menyscabant la seva fama o atemptant contra la seva pròpia estima. En aquest sentit, l'article **209 CP** castiga amb penes de multa de sis a catorze mesos aquelles injúries greus fetes amb publicitat i, en cas contrari, amb penes de tres a sis mesos. Cal destacar, però, que únicament seran constitutives de delicte aquelles injuries que, per la seva naturalesa, efectes i circumstancies, siguin tingudes en el concepte públic com a greus. A això cal afegir que les injuries que consisteixin en la imputació d'algun fet, no es consideraran com a greus a no ser que es realitzin amb coneixement de la seva falsedat o mostrant un temerari menyspreu envers la veritat.

Respecte el delicte de calúmnies, recollit a l'article **205 CP**, aquest especifica que és calumnia tota aquella imputació d'un delicte a una tercera persona amb el coneixement de la seva falsedat o amb temerari menyspreu cap a la veritat. Al seu torn, l'article **206 CP**, castiga amb penes de presó de sis mesos a dos any o multa de dotze a 24 mesos si les calumnies es propaguen amb publicitat, i amb multa de sis a dotze mesos en cas contrari. En aquest cas, per tal que es pugui aplicar el delicte de calúmnies, és condició indispensable que el fet que s'imputi, coneixent-ne la seva falsedat o amb un clar menyspreu envers la veritat, sigui un delicte tipificat al codi penal. Així ho estableix la jurisprudència del Tribunal Suprem en la sentència **STS 2029/2022**, de 25 de maig, en la que estableix que "*la imputació d'un fet delictiu cert o veritable no seria constitutiu d'un delicte de calúmnia, sens perjudici de la possibilitat que fos un delicte d'injúries*"[15]. Tampoc és considera calúmnia emprar expressions indeterminades o que suposin un greu retret social però que, en realitat, no consisteixin en imputar a una persona l'autoria d'un fet delictiu concret. En aquesta línia s'ha expressat la jurisprudència del Tribunal Suprem en l'auto **ATS 10691/2022**, de 5 de juliol, on es deixa clar que "*per integrar el delicte de calúmnia no són suficients imputacions genèriques. És essencial que siguin tan concretes i terminants que, en allò bàsic, continguin els elements requerits per definir el delicte atribuït. Per això no és calúmnia, en principi, trucar a una altra persona "estafador" o*

[15] https://www.poderjudicial.es/search/AN/openDocument/6d7ce4d86cebe7c9/20220606

"lladre" si no se li atribueixen específicament fets que siguin constitutius d'aquestes figures penals, sense perjudici que puguem estar davant d'unes injúries"[16].

D'altra banda, com ja s'ha comentat anteriorment, no és objectiu d'aquest treball entrar en el debat sobre la línia que separa entre els delictes d'injúries i calúmnies i el dret a la llibertat d'expressió consagrada a l'article **20.1.a CE**. Dit això, hi ha juristes que afirmen que voler emprar els delictes d'injúries i/o calúmnies per posar fre a les *fake news* propagades a través del ciberespai, pot tenir una utilitat molt limitada davant el reconeixement, extensió i intensitat que la Constitució Espanyola dona al dret a la llibertat d'expressió (Fernández Entralgo, 2020).

2.1.6. Delictes contra la salut pública, estafa i intrusisme

En aquest cas, l'informe de la Secretaria Tècnica de la FGE aglutina en un únic punt els delictes contra la salut pública previstos als articles **359 i seg. CP**, d'estafa dels articles **248 i seg. CP**, i intrusisme de l'article **403 CP**. El motiu per qual aquests delictes s'agrupen està directament relacionat al moment en el qual la Secretaria Tècnica de la FGE va realitzar el seu informe, en plena crisi sanitària de la COVID-19, arrel de la qual van aparèixer una gran quantitat de *fake news* escampades a través d'Internet i les TIC, sobretot mitjançant l'ús de pàgines web i xarxes socials, elaborades per falsos professionals i/o experts sanitaris, però també per aquells que realment ho eren, en les que, en ocasions amb un interès econòmic, oferien solucions o medicaments falsos o no existents per, presumptament, lluitar contra la pandèmia del coronavirus. Cal remarcar, doncs, que aquests delictes es podrien cometre de forma separada o conjunta, cosa que comportaria, en el darrer cas, un concurs ideal de delictes.

Començant amb els delictes contra la salut pública, recollits al **Capítol III del Títol XVII CP**, ens trobem, primerament, el delicte de l'article **359 CP**, que castiga amb pena de presó de sis mesos a tres anys, multa de sis a dotze mesos i inhabilitació especial per la professió o ofici de sis mesos a dos anys aquell qui, sense estar degudament autoritzat, elabori, dispensi, subministri o comercialitzi substancies nocives per la salut o productes químics que puguin causar estralls. Seguidament, l'article **360 CP** castiga amb penes de multa de sis a dotze mesos i inhabilitació especial per la professió o ofici de sis

[16] https://www.poderjudicial.es/search/AN/openDocument/f8a4eb69f9bf7bb3a0a8778d75e36f0d/20220718

mesos a dos anys aquell qui, estant autoritzat per al tràfic de substàncies o productes referides a l'article **359 CP**, els dispensi o subministri sense complir les formalitats previstes en la Llei. Al seu torn, l'article **362 bis CP** castiga amb penes de presó de sis mesos a quatre anys, multa de sis a divuit mesos i inhabilitació especial per professió o ofici d'un a tres anys aquell qui, amb coneixement de la seva falsedat o alteració, entre altres, anunciï, faci publicitat, ofereixi, vengui, faciliti, expedeixi, distribueixi o posi en el mercat, medicaments, substàncies actives, excipients, productes sanitaris, accessoris, elements o materials dels referits a l'article **362 CP**.

Si relacionem aquests delictes amb la propagació de *fake news* a través d'Internet i les TIC, cal destacar que, per tal que puguin ser penades per algun dels delictes tipificats als articles **359 i seg. CP**, és necessari que aquestes incideixin de forma negativa en la salut de les persones. Per tant, en el cas que la difusió d'una *fake news* no tingués efectes lesius envers la salut i, amb això, no posés en risc la vida o la salut de les persones, no es podria considerar com una conducta constituent d'un delicte contra la salut pública.

Seguidament, la Secretaria Tècnica de la FGE tracta els delictes d'estafa, ja que, a banda de poder esdevenir un risc per la salut, si la publicació d'una *fake news* atribuint propietats medicinals a una substancia que pogués posar en risc la vida o la salut de les persones, estigués acompanyada d'un negoci que comportés un benefici econòmic a qui la divulgués, aquesta actitud podria ser, a més, constitutiva d'un dels delicte d'estafa regulats a la **Secció 1a del Capítol VI del Títol XIII CP**. D'altra banda, com ja s'ha comentat anteriorment, es podria donar el cas que la divulgació d'una *fake news* no pogués ser considerada com un delicte contra la salut pública per no posar en perill o risc la vida o la salut de les persones, però que, de totes formes, encara pogués ser constitutiva delicte d'estafa. Un exemple seria publicar per les xarxes socials, amb l'objectiu d'obtenir un benefici econòmic, les propietats curatives de suposats medicaments, substàncies actives, excipients, productes sanitaris, accessoris, elements o materials dels referits a l'article **362 CP** que no complissin amb la funció anunciada però que, alhora, resultessin innocus per la salut. En aquest cas no seria d'aplicació cap dels delictes contra la salut pública recollits al **Capítol III del Títol XVII CP** però, en canvi, si es podria aplicar l'article **248 CP**, que considera reus d'estafa qui, amb ànim del lucre, utilitzin engany suficient per a produir error en un altre, induint-lo a realitzar un acte de disposició en perjudici propi o aliè. La pena imposada aniria de sis mesos a tres anys de presó en cas

que la quantia del defraudat superés els quatre-cents euros i, en cas contrari, s'imposaria una pena de multa d'un a tres mesos.

Finalment, la Secretaria Tècnica de la FGE introdueix en aquest apartat el delicte d'intrusisme tipificat a l'article **403 CP**, on les conductes que descriu poden entrar en concurs amb els delictes contra la salut pública i estafa descrits anteriorment. Així doncs, l'article **403 CP** castiga amb penes de multa de dotze a vint-i-quatre mesos aquell qui exercís actes propis d'una professió sense posseir el corresponent títol acadèmic expedit o reconegut a Espanya, i penes de multa de sis a dotze mesos en el cas que l'activitat professional exigís un títol oficial de capacitació i habilitació per al seu exercici. Aquestes penes es veurien transformades a penes de presó de sis mesos a dos anys si, a més, la persona s'atribuís públicament la qualitat de professional emparada pel títol referit o si exercís els actes descrits en un local o establiment obert al públic en el qual anunciés la prestació de serveis propis d'aquella professió.

Així doncs, com ja s'ha comentat anteriorment, la divulgació de *fake news* en el context de la pandèmia de la COVID-19 esdevé un clar exemple de la dimensió i l'impacte que les *fake news* poden tenir en la societat, en tant que va ser necessari que en múltiples ocasions les autoritats sanitàries i col·legis de metges haguessin de sortir a desmentir *fake news* relacionades amb suposats remeis com, per exemple, el diòxid de clor o CDS que no només resultaven inefectius contra la COVID-19, sinó que, a més, podien suposar un greu perill per la salut de les persones. També va ser necessari realitzar campanyes informatives per desmentir *fake news* en les quals s'oferien falses vacunes de la COVID-19 pagant entre tres-cents i cinc-cents euros. Cal destacar, finalment, que, amb diferència, el mitjà més emprat per la divulgació d'aquestes fake news va ser el ciberespai i, més concretament, les grans plataformes d'internet, xarxes socials i aplicacions de missatgeria instantània.

2.1.7. Delictes contra el mercat i els consumidors

El darrer punt de l'informe de la Secretaria Tècnica de la FGE estan enumerats els delictes contra el mercat i els consumidors de l'article **282 CP** i l'article **284.1.2 CP**. Amb això, el primer castiga amb pena de presó de sis mesos a un any o pena de dotze a catorze mesos de multa aquells fabricants o comerciants que, en les seves ofertes o publicitat de

productes o serveis, facin al·legacions falses o manifestin característiques incertes sobre els mateixos, de forma que puguin causar un perjudici greu i manifest als consumidors.

En aquest cas, doncs, les *fake news* divulgades a través d'Internet i les TIC esdevindran un mitjà per donar publicitat productes o serveis amb falsedat de tal forma que, per tal que sigui aplicable aquest tipus penal, es pugui causar un perjudici greu i manifest als consumidors. A més, en tant que només el poden cometre fabricants o comerciants, aquest esdevé un delicte especial propi, tal i com queda reflectit a la jurisprudència del Tribunal Suprem en la sentència **STS 1908/2004**, de 19 de març, en la que estableix que els elements constitutius del delicte de l'article **282 CP** són "*1r. Subjecte actiu ha de ser un fabricant o comerciant. Es tracta, doncs, d'un delicte especial propi (...)*"[17]. Que es tracti d'un delicte d'aquestes característiques limita molt l'àmbit d'aplicació d'aquest tipus penal ja que en aquest tipus de delictes la condició personal de l'autor és la que fonamenta la pena.

Per la seva banda, l'article **284.1.2 CP**, encara resulta més específic i castiga amb pena de presó de sis mesos a sis anys, multa de dos a cinc anys o del triple del benefici obtingut o dels perjudicis evitats, si la quantitat fos més elevada, i inhabilitació especial per un temps de dos a cinc anys qui de forma directa o indirecta o per mitjà d'un medi de comunicació, Internet o les TIC, o per qualsevol altre mitjà, difonguin noticies o rumors sobre persones o empreses, oferint de propòsit dades econòmiques total o parcialment falsos amb la fi d'alterar o preservar el preu d'un instrument financer o manipular el càlcul d'un índex de referència quan, d'aquesta pràctica, obtingués, per si mateix o un tercer, un benefici superior a dos-cents cinquanta mil euros o es causés un perjudici d'idèntica quantitat, que l'import dels fons empleats fos superior a dos milions d'euros o que es causés un greu impacte en la integritat del mercat.

Així doncs, amb la inclusió del delicte de l'article **284.1.2 CP**, la Secretaria Tècnica de la FGE busca protegir el correcte funcionament dels mercats financers contra aquelles *fake news* que busquin alterar el preu de cotització d'un valor o instrument financer. En aquest sentit, en tractar-se d'un delicte especial propi que exigeixen que es compleixin uns requisits molt específics, per tal de poder aplicar-lo, serà necessari

[17] https://www.poderjudicial.es/search/AN/openDocument/5fdbe682ea59328c/20040503

realitzar un estudi exhaustiu de les circumstàncies que concorrin i dels elements dels tipus penals (Rodríguez Gutiérrez, 2020). Tot i això, un correu electrònic enviat de forma massiva amb declaracions suposadament atribuïdes a Doug McMillon, conseller delegat de *Walmart*, on afirmava que emprarien litecoin com a mitjà de pagament i, de forma similar, publicacions en diverses webs i xarxes socials anunciant que *Amazon* agregaria el bitcoin com a mitjà de pagament, van provocar que el preu d'aquestes criptomonedes augmentés de forma considerable el seu valor de mercat fins que les esmentades informacions van ser desmentides, esdevenen un clar exemple de *fake news* creades i divulgades pel ciberespai amb l'objectiu d'alterar el preu d'un instrument financer, en aquest cas les criptomonedes bitcoin i litecoin. En aquests casos, doncs, podria ser d'aplicació el delicte de l'article **284.1.2 CP**.

2.2. Lluita legal contra les *fake news* a Europa

Per la Unió Europea, la lluita contra les *fake news* divulgades a través del ciberespai, juntament amb la realitzada contra els discursos d'odi i discriminació, així com l'amenaça que poden representar els ciberatacs, ha esdevingut un dels reptes més importants d'afrontar i que, alhora, ha generat més debat a la seu del Parlament Europeu. Així ha quedat demostrat en els diversos plens realitzats, on els membres el Parlament han mostrat diverses postures allunyades i enfrontades entre si.

Amb això, cal remarcar que, a nivell Europeu, s'ha decidit utilitzar més el terme desinformació que *fake news*, en tant es considera que l'expressió *fake news* resulta inadequada per captar la complexitat que comporta problema de la desinformació i, a més, pot resultar enganyosa ja que alguns polítics i els seus partidaris se l'han apropiat per senyalar i descartar aquella cobertura informativa que consideren desagradable o no afí a la seva ideologia política. Així doncs, la desinformació es defineix com aquella "*informació falsa, inexacta o enganyosa dissenyada, presentada i promoguda per causar danys públics intencionadament o amb ànim de lucre, que es pot adreçar específicament a diversos sectors, com la salut, la ciència, l'educació i les finances, entre altres*", afegint que "*l'impuls per la producció i promoció de la desinformació té l'objectiu d'obtenir*

guanys econòmics o aconseguir objectius polítics o ideològics, i es pot agreujar segons com els diferents públics i comunitats la reben, participen i amplifiquen"[18].

Amb això, la reunió del Consell Europeu en el qual es van publicar les Conclusions dels dies 19 i 20 de març de 2015[19] esdevé un punt de partida a Europa en la lluita contra les *fake news*. Concretament, el punt número tretze d'aquestes Conclusions apuntava que "*El Consell Europeu ha destacat la necessitat de contrarestar les actuals campanyes de desinformació de Rússia i ha convidat a l'alta representant a que, en cooperació amb els Estats membres i les institucions de la UE, preparin abans de juny un pla d'acció sobre comunicació estratègica. La creació d'un equip de comunicació constitueix una primera mesura en aquest sentit"*.

A partir d'aquest moment, des del Parlament Europeu i la Comissió Europea s'han engegat diverses d'iniciatives per lluitar contra les *fake news*, entre les quals ens trobem, per nombrar-ne algunes, amb el **Pla d'acció contra la desinformació**[20], presentat el 5 de desembre de 2018, el qual es basa en quatre grans pilars per donar una resposta coordinada a la desinformació: "*la millora de la capacitat de les institucions de la Unió per detectar, analitzar i exposar la desinformació; el reforç de les respostes coordinades i conjuntes a la desinformació; la mobilització del sector privat per combatre la desinformació; i, finalment, l'augment de la sensibilització i la capacitat de resposta de la societat*". Aquest pla contempla la creació d'un **Sistema d'Alerta Ràpida o RAS**[21] (sigles en anglès de Rapid Alert System), activat el 19 de març de 2019, per informar instantàniament sobre campanyes de desinformació, intercanviar dades i prendre les decisions oportunes entre els estats membres. Destacar que Espanya participa en el **RAS** mitjançant la **Comissió Permanent contra la Desinformació**[22]. També cal esmentar el **Pla d'acció per a la democràcia europea: reforçar les democràcies de la UE**[23], presentat el 3 de desembre de 2020, el qual estableix una sèrie de mesures en base a tres grans pilars: "*la promoció d'unes eleccions lliures i justes; el reforç de la llibertat dels mitjans de comunicació; i la lluita contra la desinformació*". En relació a les *fake news*,

[18] https://www.ami.info/wp-content/uploads/2018/03/HLEGReportonFakeNewsandOnlineDisinformation.pdf
[19] https://data.consilium.europa.eu/doc/document/ST-11-2015-INIT/es/pdf
[20] https://data.consilium.europa.eu/doc/document/ST-15431-2018-INIT/es/pdf
[21] https://www.eeas.europa.eu/sites/default/files/ras_factsheet_march_2019_0.pdf
[22] https://www.boe.es/buscar/doc.php?id=BOE-A-2020-13663
[23] https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52020DC0790&from=ES

proposa perfeccionar els instruments emprats per la Unió Europea per contrarestar les ingerències estrangeres i lluitar-hi de forma eficaç, així com el disseny i la implantació de nous que permetin imposar sancions als responsables de la seva creació i difusió.

Un altre exponent va ser la creació d'una **Comissió Especial sobre Ingerències Estrangeres en Tots els Processos Democràtics de la Unió Europea, en particular la Desinformació** o **INGE**[24] (sigles en anglès de Special Committee on Foreign Interference in all Democratic Processes in the European Union, including Disinformation), al qual el Parlament Europeu va encomanar la formulació d'un enfocament a llarg termini per fer front a les proves d'ingerències estrangeres de les institucions i processos democràtics de la Unió Europea i els seus estats membres. El 8 de febrer de 2022 l'**INGE** va presentar un informe[25] en el qual, entre altres punts, es destaca la necessitat de reforçar la resiliència davant les *fake news* o les campanyes de desinformació mitjançant "*la consciència situacional, l'alfabetització mediàtica i informativa, el pluralisme dels mitjans de comunicació, el periodisme independent i l'educació*".

En aquest context, les dues grans iniciatives que des del estaments de la Unió Europea s'han endegat per tal de lluitar contra les *fake news*, i les que, per tant, seran les que s'estudien en el present treball, són el **Codi de bones pràctiques de la Unió en matèria de desinformació**, impulsat per la Comissió Europea i que ja s'esmenta al **Pla d'acció contra la desinformació**, i la **Llei de Serveis Digitals**, que va ser aprovada el passat 16 de novembre de 2022. Ambdós han resultat cabdals en tant que han suposat l'entrada en vigor d'un marc de regulació i rendició de comptes de les plataformes en línia, entre altres, en el relacionat amb les *fake news* i les campanyes de desinformació a divulgades través d'Internet i les TIC.

2.2.1. El Codi de bones pràctiques de la Unió en matèria de desinformació

La primera versió d'aquest **Codi** va ser publicada el 26 de setembre de 2018[26] i, tot no tractar-se d'una norma legal, es considera adient tractar-lo ja que va esdevenir el primer marc mundial d'aquest tipus, en el qual grans plataformes d'Internet, empreses punteres del sector tecnològic i agents del sector de la publicitat, van acordar de forma

[24] https://www.europarl.europa.eu/committees/es/inge/home/highlights
[25] https://www.europarl.europa.eu/doceo/document/A-9-2022-0022_ES.pdf
[26] https://digital-strategy.ec.europa.eu/es/library/2018-code-practice-disinformation

voluntària un codi de bones pràctiques i normes d'autoregulació per fer front i combatre les campanyes de desinformació en línia o *fake news*.

Aquest **Codi**, al qual inicialment s'hi van adherir grans plataformes d'Internet com *Facebook, Google, Twitter* i *Mozilla*, i a les que posteriorment van seguir *Microsoft* (any 2019) i *TikTok* (any 2020), proposava un total d'onze objectius que les entitats adherides establien com estratègiques per afrontar els reptes associats a la desinformació, entre els destaquen la inclusió d'elements per lluitar contra la desinformació; reduir la visibilitat de la desinformació; reduir els ingressos als proveïdors de desinformació; intensificar la lluita i l'eficàcia per tancar comptes falsos i establir sistemes i regles de marcatge clars per ordinadors zombis; i adoptar les mesures oportunes per permetre una accés compatible amb la protecció de la intimitat a l'hora de dur a terme activitats de verificació de fets i recerca.

Relacionats amb aquestes estratègies, el **Codi** establia un total de vint-i-un compromisos agrupats en cinc àmbits diferents: l'***anàlisi de la col·locació dels anuncis***, la ***publicitat política i publicitat sobre aspectes concrets***, la ***integritat dels serveis***, la ***capacitació dels consumidors***, i ***la capacitació de la comunitat investigadora***. A banda, els signataris del **Codi** es comprometien una avaluació i seguiment d'aquest, fet que implicava, entre altres punts, a redactar un informe anual sobre el treball dut a terme per tal de lluitar contra la desinformació. A més, s'instaurava la necessitat de dur a terme, amb una periodicitat també anual, reunions en les quals es revisaria el **Codi** i s'adoptarien les mesures addicionals en cas que es considerés necessari.

En aquest context, el setembre de l'any 2020, la Comissió Europea va publicar una avaluació del **Codi**[27], on destacava que aquest havia esdevingut un marc valuós a partir del qual s'havia pogut establir un diàleg directe i estructurat amb els seus signataris, i on es garantia una transparència i rendició de comptes més grans en relació a les seves polítiques en matèria de desinformació, tal com es va poder demostrar en la seva utilitat per supervisar, limitar i contrarestar l'impacte la desinformació relacionada amb la

[27] https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement

pandèmia de la COVID-19[28], en el transcurs de la qual les plataformes *Facebook*, *Twitter*, *Google*, *Microsoft*, *TikTok* i *Mozilla* van realitzar un conjunt d'informes per tal de constatar i garantir els seus esforços en la lluita per limitar la desinformació generada arrel de la pandèmia. Tanmateix, aquesta avaluació també va mostrar l'existència d'una sèrie de llacunes i deficiències importants i, per això, el maig de l'any 2021, es van publicar unes orientacions[29] per tal de reforçar el **Codi**.

Tot això va portar a que el 16 de juny 2022 es presentés una nova versió molt més àmplia i reforçada del **Codi**[30], el qual va ser referendat per trenta-quatre signataris. En aquesta nova versió, s'estableixen mesures i compromisos més ambiciosos per lluitar contra la desinformació en línia. Concretament el nou **Codi** conté quaranta-quatre compromisos i cent-vint-i-vuit mesures específiques, agrupats en set àmbits diferents: la ***reducció dels incentius financers per als proveïdors de desinformació***, la ***transparència de la publicitat política***, *el* ***garantir la integritat dels serveis***, l'***empoderament dels usuaris***, *l'****empoderament dels investigadors***, l'***empoderament de la comunitat de verificació de fets***, el ***centre de transparència i grup de treball*** i, finalment, l'***enfortiment del marc de supervisió***. Tanmateix, cal recordar en aquest punt que l'adhesió al **Codi** és voluntària i, per tant, les gran plataformes d'Internet signatàries poden decidir en tot moment el seu nivell de compromís amb aquest, és a dir; decidir a què es comprometen i a què no. Un exemple el trobem amb l'anunci realitzat pel Comissari europeu de Mercat Interior i Serveis, Thierry Breton, el qual, el 26 de maig de 2023, va publicar al seu perfil de *Twitter* que aquesta xarxa social havia decidit abandonar el **Codi**[31].

Seguint amb les novetats introduïdes en el nou **Codi** reforçat, cal destacar que aquest contempla obligacions de transparència pels sistemes d'intel·ligència artificial per tal que s'identifiqui i etiqueti el contingut creat amb aquesta tecnologia, així qualsevol usuari pot tenir la capacitat d'identificar un text o una imatge que no hagin estat produïts per persones reals. Amb això, el **Codi** estableix que "*els signants rellevants establiran o confirmaran les seves polítiques vigents per contrarestar les pràctiques manipuladores prohibides per als sistemes d'IA que generen o manipulen contingut, com ara avisar els*

[28] https://digital-strategy.ec.europa.eu/en/library/first-baseline-reports-fighting-covid-19-disinformation-monitoring-programme
[29] https://digital-strategy.ec.europa.eu/en/library/guidance-strengthening-code-practice-disinformation
[30] https://digital-strategy.ec.europa.eu/es/library/2022-strengthened-code-practice-disinformation
[31] https://twitter.com/ThierryBreton/status/1662194595755704321

usuaris i detectar aquest contingut de manera proactiva". En aquest sentit, el 14 de juny de 2023 el Parlament Europeu va adoptar una posició negociadora[32] sobre una nova **Llei d'Intel·ligència Artificial** (**AI Act**)[33] que ha de definir els límits d'aquesta tecnologia. A l'espera que la Comissió Europea elabori la proposta de llei i aquesta sigui aprovada, previst per l'any 2026, actualment el nou **Codi** reforçat és l'única eina existent a nivell Europeu que contempla la lluita contra les *fake news* i la desinformació creades a partir de d'intel·ligència artificial.

Una altra de les novetats introduïdes és l'ampliació dels agents involucrats, afegint als anteriors, els verificadors de fets, les plataformes emergents o especialitzades, la societat civil i aquelles organitzacions amb coneixements específics en matèria de desinformació. Dels àmbits i agents involucrats en el nou codi reforçat, cal destacar la figura de la comunitat de **verificadors de fets** o **fact-checkers**. En aquest sentit, el signants del **Codi**, en el marc d'una estratègia eficaç per lluitar contra la desinformació, es comprometen "*a establir un marc de transparència, cooperació estructurada, oberta, econòmicament sostenible i no discriminatòria entre ells i la comunitat de verificació de fets de la Unió Europea pel que fa als recursos i el suport posats a disposició d'aquests*", "*a integrar, mostrar o utilitzar de manera coherent el treball dels verificadors de fets en els serveis, processos i continguts de les seves plataformes; amb cobertura total a tots els estats membres i llengües*", "*a proporcionar als verificadors de fets un accés ràpid i, sempre que sigui possible, automatitzat a la informació que sigui pertinent per ajudar-los a maximitzar la qualitat i l'impacte de la verificació de fets*", i "*a operar (les organitzacions de verificació de fets) sobre la base de normes ètiques i de transparència estrictes i a protegir la seva independència*".

Finalment, el **Codi** reconeix la importància que, per tal que els **verificadors de fets** puguin ser realment efectius en la lluita contra la desinformació, és necessari que siguin, de manera verificable, independents de qualsevol institució de caire partidista i transparents en les seves finances, organització i metodologia. També se'ls requereix que estiguin dedicats de manera coherent i contínua a la verificació de fets, ja sigui com a

[32] https://www.europarl.europa.eu/news/es/headlines/society/20230601STO93804/ley-de-ia-de-la-ue-primera-normativa-sobre-inteligencia-artificial

[33] https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence

signants verificats del codi de principis de la **Xarxa Internacional de Verificació de Fets** o **IFCN** (sigles en anglès de la International Fact-Checking Network), com a membres de la xarxa de verificadors de fets de l'**Observatori Europeu de Mitjans Digitals** o **EDMO** (sigles en Anglès d'European Digital Media Observatory), o del que en aquell moment es va anomenar com el futur Codi d'Integritat Professional per Organitzacions Europees Independents de verificació de fets, i que posteriorment va acabar generant en el **Codi Europeu d'Estàndards per a Organitzacions Independents de Verificació de Fets**, que esdevé de necessari compliment per poder formar part de la **Xarxa Europea d'Estàndards de Verificació de Fets** o **EFCSN** (sigles en anglès d'European Fact-Checking Standards Networks).

2.2.2. Reglament (UE) 2022/2065 relatiu a un mercat únic de serveis digitals o Llei de Serveis Digitals

A diferència de la normativa espanyola, la **Llei de Serveis Digitals** o **DSA**[34] (sigles en anglès de Digital Services Act) està pensada per ser aplicada a tots aquells serveis digitals que connecten els consumidors amb béns, serveis i continguts proporcionats a través del ciberespai. A grans trets, estableix noves responsabilitats per aquests serveis digitals, esdevé un reforç de les garanties dels drets fonamentals en línia, i atorga noves competències de supervisió pública de la Comissió Europea a les plataformes en línia, tant a escala nacional com de la Unió Europea. En aquest sentit, la **DSA**, al tractar-se d'una norma europea amb rang de Reglament[35], esdevé d'aplicació directa a tots els països que formen part de la Unió Europea.

La Comissió Europea va presentar la seva primera proposta sobre la **DSA** el 15 de desembre de l'any 2020, i la va basar en la **Directiva 2000/31/CE**[36], de 8 de juny de 2000, sobre el comerç electrònic per fer front als nous reptes en línia, pedra angular a la Unió Europea en matèria sobre regulació en l'àmbit digital. Amb això, després de superar el corresponent tràmit parlamentari, el 27 d'octubre de l'any 2022 la **DSA** va ser publicada al Diari Oficial de la Unió Europea i, tal i com estipula el seu **article 93**, va entrar en vigor el 16 de novembre del mateix any i va fixar com a data per la seva aplicació efectiva

[34] https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32022R2065&from=EN
[35] https://european-union.europa.eu/institutions-law-budget/law/types-legislation_es
[36] https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32000L0031&from=ES

el dia 17 de febrer de l'any 2024. Tot i això, el mateix article enumera una sèrie d'articles que havien de ser d'aplicació des del mateix dia d'entrada en vigor de la norma.

Seguint el cronograma d'implantació, l'**article 24.2** estableix com a termini màxim dia el 17 de febrer de 2023 per tal que les plataformes en línia o motors de cerca publiquessin el seu nombre d'usuaris. Totes aquelles superessin els quaranta-cinc milions d'usuaris, l'equivalent actual al 10% de la població europea, serien designades per la Comissió Europea com a plataforma en línia de mida molt gran o **VLOP** (sigles en anglès de Very Large Online Platform) o motor de cerca en línia de mida molt gran o **VLOSE** (sigles en anglès de Very Large Online Search Engine). El 25 d'abril de 2023 la Comissió Europea va publicar un primer llistat amb un total de dinou companyies, disset designades com a **VLOP**: *Alibaba AliExpress*, *Amazon Store*, *Apple AppStore*, *Booking.com*, *Facebook*, *Google Play*, *Google Maps*, *Google Shopping*, *Instagram*, *LinkedIn*, *Pinterest*, *Snapchat*, *TikTok*, *Twitter*, *Wikipedia*, *YouTube* i *Zalando*, i dos designades com a **VLOSE**: *Bing* i *Google Search*[37].

Les **VLOP** i les **VLOSE**, tal i com s'especifica al **Considerant 76**, poden comportar riscos sistèmics per la societat i, en conseqüència, han de complir unes obligacions de diligència deguda més exigents, proporcionals en relació amb el seu impacte social. En relació a les *fake news*, el **Considerant 84** estableix que, en l'avaluar els riscos sistèmics "*aquests prestadors també s'han de centrar en la informació que no sigui il·lícita, però que contribueixi als riscos sistèmics identificats en aquest Reglament. Per tant, aquests prestadors han de dedicar especial atenció a com s'utilitzen els seus serveis per difondre o amplificar continguts incorrectes o enganyosos, inclosa la desinformació"*, i afegeix que "*aquests riscos poden derivar-se, per exemple, de l'ús no autèntic del servei, com ara la creació de comptes falsos, l'ús de bots o l'ús enganyós d'un servei, i altres comportaments totalment o parcialment automatitzats, que poden donar lloc a la difusió ràpida i estesa d'informació al públic que sigui un contingut il·lícit o incompatible amb les condicions generals d'una plataforma online o un motor de cerca en línia i que contribueixi a campanyes de desinformació."*. Per aquests motius, l'**article 33.6** especifica que, des del moment en el qual són designades com a tals, les **VLOP** i les **VLOSE** únicament disposen de quatre mesos per aplicar la **DSA** i complir amb les

[37] https://digital-strategy.ec.europa.eu/es/policies/dsa-vlops

obligacions addicionals establertes per aquesta, cosa que inclou realitzar i proporcionar a la Comissió Europea la seva primera avaluació anual de riscos i realitzar aquesta avaluació almenys un cop a l'any (**article 34.1**). Després de publicar el primer llistat, la Comissió Europea ha establert com a data el 25 d'agost de 2023 perquè les dinou companyies anunciades com a **VLOP** i **VLOSE** comencin a complir amb les noves obligacions imposades per la **DSA**.

Abans de la data general d'entrada en vigor de la **DSA**, el 17 de febrer de 2024, és necessari que els Estats membres de la Unió Europea designin una o diverses autoritats competents responsables de la supervisió dels prestadors dels serveis intermediaris i de la correcta l'execució de la **DSA** (**article 49.1**). A més, hauran de nombrar una de les autoritats competents com a coordinador de serveis digitals el qual, de forma general, serà el responsable d'aquelles matèries relacionades amb la supervisió i la garantia del compliment de la **DSA** (**article 49.2**). Per això, la **DSA** atorga als coordinadors de serveis digitals, respecte dels prestadors de serveis intermediaris subjectes a la seva competència, les facultats d'investigació (**article 51.1**), execució (**article 51.2**) i, en cas que sigui necessari, l'aplicació de les mesures llistades a l'**article 51.3**. La **DSA** també crea la **Junta Europea de Serveis Digitals**, regulada als **articles 61**, **62** i **63**, un grup consultiu independent integrat per coordinadors de serveis digitals encarregada de la supervisió dels prestadors de serveis intermediaris.

Un cop detallada la cronologia d'implantació de la **DSA**, cal especificar que aquesta s'organitza en cinc grans capítols, dels quals resulten d'interès el **capítol II**, que regula la responsabilitat dels prestadors de serveis intermediaris, el **capítol III**, on es detallen les obligacions aplicables a tots els prestadors de serveis en línia per crear un entorn en línia transparent i segur, i, finalment, el **capítol IV**, on es desenvolupa el marc d'aplicació, cooperació, imposició de sancions així com la seva execució. Si ens centrem en el relacionat amb les *fake news*, un dels objectius cabdals de la **DSA** és facilitar l'eliminació de productes, serveis o continguts il·lícits, així com desinformació divulgats a través del ciberespai. Per això, la **DSA** crea noves i exhaustives obligacions per tal que les plataformes en línia redueixin els possibles danys causats i contrarestin els riscos en línia, introduint una reforçada protecció dels drets dels usuaris. A més, proporciona un conjunt de normes enfocades als serveis d'intermediació en línia, sobretot aquells classificats com a **VLOP** i **VLOSE**, sobre la forma en la que han de dissenyar els seus

serveis i procediments, i també els imposa noves responsabilitats enfocades a limitar la difusió a través de les noves tecnologies, Internet i les TIC de productes o continguts il·lícits i/o desinformació.

És per això que la **DSA** desenvolupa mesures per lluitar-hi i estableix, per una banda, la necessitat d'implantar nous mecanismes que permetin als usuaris assenyalar-los i, per l'altra, l'obligació per part de les plataformes de cooperar amb els anomenats "alertants fiables" especialitzats (els **verificadors de fets o fact-checkers**) per detectar i eliminar aquests tipus de continguts. Tot això amb la intenció de reduir els riscos sistèmics relacionats amb la manipulació i/o la desinformació en línia. En aquest sentit, al **Considerant 9** especifica que la **DSA** "*harmonitza plenament les normes aplicables als serveis intermediaris al mercat interior amb l'objectiu de garantir un entorn en línia segur, predictible i digne de confiança i aborda la difusió de continguts il·lícits en línia i els riscos per a la societat que pot generar la difusió de desinformació o altres continguts*".

Tot i això, cal destacar que la **DSA** manté l'exempció de responsabilitat per als intermediaris en línia, en els quals s'inclouen les plataformes en línia i els motors de cerca, per la qual no esdevenen responsables dels comportaments il·lícits realitzats pels seus usuaris, a no ser que tinguin coneixement de la seva il·legalitat o il·licitud i no actuïn amb la diligència deguda per tal d'impedir-los i/o eliminar-los, cosa que també inclou tot el relacionat amb la divulgació de desinformació. En aquest darrer cas, la **DSA** preveu fortes penes de multa que poden arribar fins al 6% del seu volum de facturació mundial en l'exercici fiscal anterior (**article 52.3**). A més, la **DSA** pretén garantir que sigui possible actuar de forma ràpida i efectiva contra els continguts, béns o serveis il·lícits, però també que ni les plataformes en línia ni els motors de cerca no rebin incentius per eliminar continguts legítims i tampoc es vegin pressionades a controlar els seus usuaris.

És per tot això que la **DSA** fomenta l'adhesió dels prestadors de serveis i intermediaris en línia en els diversos codis de conducta de la Unió Europea. Així ho especifica al **Considerant 104**, on destaca que "*cal explorar mesures de reducció de riscos relatives a tipus concrets de continguts il·lícits a través d'acords d'autoregulació i corregulació*.", entre els quals, tal i com detalla el **Considerant 106**, hi ha el **Codi de bones pràctiques de la Unió en matèria de desinformació** tractat a l'apartat anterior.

3. LLUITA TECNOLÒGICA CONTRA LES FAKE NEWS

La lluita tecnològica contra la creació i divulgació de *fake news* al ciberespai, el segon gran pilar de la batalla contra les *fake news*, es pot enfocar des de diverses perspectives: una més tècnica, en la que es requereix tenir coneixements informàtics en la tecnologia emprada per combatre-hi, i una segona menys tècnica però més pràctica en el sentit que pot estar a l'abast de qualsevol internauta, ja que no resulta imprescindible tenir expertesa en el camp de la informàtica, sinó que és suficient amb estar familiaritzat en l'ús de les noves tecnologies, Internet i les TIC.

Si s'enfoca aquesta lluita des de la vessant més tècnica, es pot parlar de propostes com, per exemple, l'ús de la tecnologia **blockchain**. De forma molt resumida, es pot definir la **blockchain**, o "*cadena de blocs*", com "*un llibre de comptabilitat incorruptible de transaccions econòmiques que es pot programar no només per realitzar transaccions financeres, sinó per registrar pràcticament tot allò que té valor*" (Tapscott & Tapscott, 2018). La cadena de blocs que conforma la **blockchain** no està emmagatzemada en un únic servidor, sinó que es troba distribuïda per tots els nodes de la xarxa i, tot i que qualsevol node pot sol·licitar que s'hi afegeixi una nova transacció, aquesta només serà acceptada si els usuaris del sistema en validen la seva legitimitat. Un cop verificada, s'emmagatzemarà emprant un sistema d'encriptació de tal manera que, un cop els paquets s'incorporen a la cadena de blocs no podran ser eliminats per cap usuari i resultarà gairebé impossible manipular les dades que continguin. Per tant, la inviolabilitat de la **blockchain** està garantida ja que la informació no està centralitzada en un únic custodi o intermediari, sinó que es troba distribuïda entre tots els usuaris del sistema (Parrondo, 2017).

La **blockchain** esdevé, doncs, una forma segura, transparent i descentralitzada de registrar transaccions, i aquesta capacitat de registrar tot tipus de transaccions d'usuari a usuari de forma eficient, segura, verificable i immutable indica que pot ser emprada, entre moltes altres aplicacions, per combatre les *fake news*. En aquest punt és on destaquen iniciatives com "*Civil*", la qual, l'any 2018, va esdevenir la primera plataforma de **blockchain** creada i emprada amb fins periodístics, i on es buscava impulsar un periodisme independent, veraç i de qualitat, amb l'objectiu, entre altres, de lluitar contra les *fake news* i la desinformació (Sintes-Olivella, Xicoy-Comas, & Yeste-Piquer, 2020). Tot i que l'any 2020 el projecte es va cancel·lar, esdevé un clara mostra de com es poden

utilitzar els avantatges que ofereix la tecnologia **blockchain** per fomentar un periodisme de qualitat i alhora combatre contra la creació i divulgació de *fake news* en línia.

D'altra banda, si s'enfoca aquesta lluita des d'una perspectiva menys tècnica, existeixen diverses iniciatives que busquen dotar a qualsevol internauta d'aquells recursos i eines que els permetin detectar les *fake news* que es puguin trobar mentre naveguen pel ciberespai. De les diverses opcions existents, ens trobem amb els recursos proporcionats per **OSINT** (sigles en anglès de Open Source INTelligence o intel·ligència de fonts obertes). En aquests sentit, per l'Institut Nacional de Ciberseguretat d'Espanya (conegut per les sigles INCIBE[38]), els recursos **OSINT** "*fan referència al coneixement recopilat a partir de fonts d'accés públics*" on, després d'un procés de cerca, selecció i adquisició d'informació, i el posterior anàlisi d'aquesta informació, s'obté un coneixement útil i aplicable en diversos àmbits[39]. Amb això, entre les múltiples aplicacions on es poden emprar els recursos **OSINT**, n'hi ha que han estat específicament dissenyades i desenvolupades per combatre la creació i propagació de *fake news* al ciberespai, fet que resulta d'interès pel present treball.

Finalment, no es pot deixar de parlar de la lluita tecnològica sense esmentar la figura dels **verificadors de fets** o **fact-checkers** els quals, tot i no representar en si mateixos cap tipus de tecnologia, si que empren les eines que els proporcionen les noves tecnologies (entre les quals es troben els recursos **OSINT**) per lluitar contra la creació i divulgació de *fake news*, esdevenint un actor clau en aquesta lluita, tal i com així ho han reconegut organismes europeus com el Consell d'Europa i el Parlament Europeu.

En aquest sentit, sense voler menystenir la part tècnica, el present apartat es centra en la lluita que s'endega des de la part més pràctica, en tant és la que resulta més accessible per qualsevol tipus d'usuari que hi pugui estar interessat, siguin quins siguin els seus coneixements informàtics. Per això, aquest treball desgrana, per una banda, alguns dels recursos **OSINT** emprats per lluitar contra les *fake news* i, per l'altra, fa un estudi en detall de la figura dels **verificadors de fets** que, tal i com ja s'ha esmentat, han esdevingut una figura crucial per combatre la propagació de *fake news* al ciberespai.

[38] https://www.incibe.es/
[39] https://www.incibe-cert.es/blog/osint-la-informacion-es-poder

3.1. Recursos OSINT per lluitar contra les *fake news*

Com ja s'ha esmentat anteriorment, els recursos **OSINT** són totes aquelles eines i aplicacions que es nodreixen de fonts que són accessibles públicament. Aquestes fonts poden ser qualsevol tipus de dades i informació, cosa que inclou fotografies, vídeos i perfils d'usuaris, entre altres. En aquest punt, cal destacar que la característica principal dels recursos **OSINT** és que es nodreixen de dades públiques i, per tant, no vulneren el dret a la privadesa, ja que s'aprofiten de la publicitat, entre altres, de les dades que els mateixos usuaris faciliten quan interactuen en el ciberespai (Toro-Alvarez, Jaimes, & Bonilla-Duitama, 2018). Un altre factor important és que els recursos **OSINT** s'ofereixen de forma gratuïta a qualsevol usuari que els vulgui emprar i, a més, en moltes ocasions són de codi obert, és a dir; són aplicacions que faciliten el seu codi font per tal que aquell qui hi estigui interessat el pugui inspeccionar, modificar i millorar.

Amb això, quan es divulga una *fake news* a través del ciberespai, la majoria de vegades emprant com a mitjà les grans plataformes d'Internet, xarxes socials i aplicacions de missatgeria instantània, en moltes ocasions, a banda de la informació escrita, aquesta va acompanyada per una imatge o vídeo, amb la intenció així de donar una aparença més realista a la informació que s'està transmetent. Per tant, tenir la capacitat de verificar l'autenticitat o falsedat d'aquesta imatge o vídeo pot resultar de gran ajuda per discernir si la informació que s'està divulgant és autèntica o es tracta d'una *fake news*. A més, també resulta interessant poder esbrinar si un compte d'usuari és real, es tracta d'un compte bot o és compte emprat bàsicament per difondre *fake news*. Així doncs, analitzar el comportament dels perfils a les xarxes socials pot ajudar a respondre aquesta qüestió. Per dur a terme aquestes verificacions es poden emprar diverses tècniques, i és en aquest punt on entren en joc els recursos que **OSINT** ofereix, els quals es detallen a continuació.

3.1.1. Verificació d'imatges i vídeos

Per discernir si una imatge o un vídeo són o no són autèntics, es poden realitzar una sèrie de comprovacions per verificar la seva originalitat, determinar-ne la font, i el lloc i la data on van ser enregistrats. Per dur a terme aquestes tasques existeixen una sèrie d'eines i recursos **OSINT** que permeten respondre a les qüestions plantejades. En aquest sentit, la *recerca inversa d'imatges*, per comprovar si la imatge es autèntica o ja existia amb anterioritat; l'*anàlisi de les metadades*, que permet conèixer les dades que acompanyen la imatge o el vídeo, la *geolocalització*, que permet verificar si el lloc en el

que es pretén que va ser realitzada una imatge o vídeo es correspon realment al lloc on realment es va enregistrar i, finalment, emprar *eines per traduir imatges* per constatar si en aquestes efectivament es diu el que s'anuncia en la notícia o publicació, poden esdevenir eines de gran utilitat.

3.1.1.A. Recerca inversa d'imatges

La recerca inversa d'imatges és una tècnica molt simple que permet conèixer quin és l'origen d'una imatge que circula pel ciberespai, les pàgines web en les que apareix i, fins i tot, arribar a trobar la persona que la va crear. Amb aquesta tècnica es pot saber si una imatge anteriorment ja havia estat emprada amb altres fins o propòsits, cosa que resulta de gran ajuda a l'hora de desemmascarar una *fake news*.

Es poden trobar una gran diversitat de recursos **OSINT** que permeten realitzar aquest tipus de recerca, entre els quals hi ha les pàgines web *Microsoft Bing*[40], *Google Images*[41], *Yandex*[42] o *TinEye*[43]. Cadascuna d'aquestes ofereix a l'usuari, a partir de la imatge facilitada (ja sigui mitjançant la pròpia imatge o la seva URL), un llistat d'aquelles pàgines web que la contenen, el contingut relacionat i imatges que puguin resultar similars o que, al seu torn, continguin la pròpia imatge. Per destacar-ne un, *Yandex* ofereix a l'usuari la possibilitat d'ordenar els resultats segons l'antiguitat de la imatge, la millor coincidència, la imatge que hagi estat més modificada i la imatge que ocupi més espai.

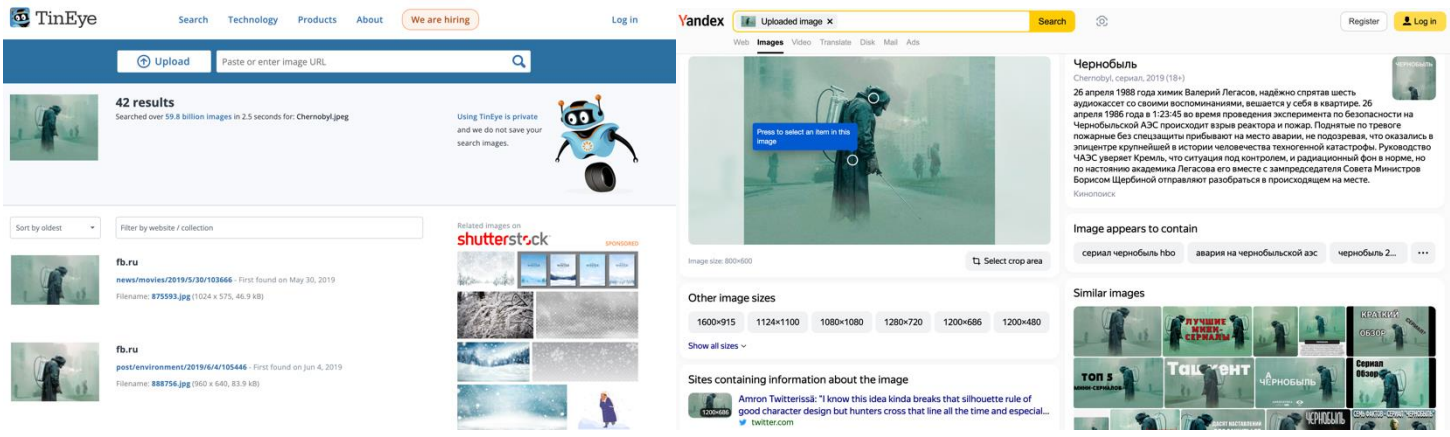


Figura 1. Recerca inversa d'imatges emprant *TinEye* i *Yandex*.

[40] https://www.bing.com/?cc=ar
[41] https://images.google.com/
[42] https://yandex.com/images/
[43] https://tineye.com/

3.1.1.B. Anàlisi de les metadades d'imatges i vídeos

Les metadades són aquelles dades que tot arxiu o fitxer digital conté que no són visibles a simple vista i que proporcionen informació sobre aquest. És per això que es coneixen com "les dades de les dades" (Quinto Huamán, Armas Vega, Sandoval Orozco, & García Villalba, 2016). Quan es tracta d'imatges o vídeos, la especificació de format d'arxiu d'imatge intercanviable o **EXIF** (sigles en anglès d'Exchangeable Image File Format), esdevé el contenidor de metadades més comú emprat pels dispositius digitals que enregistren imatges i vídeos, entre els quals es troben les càmeres digitals i els smartphones. Així doncs, les dades **EXIF** proporcionen una valuosa font d'informació, ja que contenen les metadades que el dispositiu digital omple automàticament en el moment en el qual realitza la foto o el vídeo, cosa que permet conèixer, entre altres, la data i hora de la seva creació, el seu format, les dades del dispositiu emprat, la mida, la duració (del vídeo) i, fins i tot, les coordenades geogràfiques del lloc on va ser enregistrat.

Tanmateix, no es pot afirmar de forma inequívoca que les dades obtingudes a partir de l'estudi de les metadades són del tot fiables, ja que existeixen una gran quantitat d'aplicacions que permeten modificar-les. De tota manera, en la mesura que no s'hagin vist afectades, la informació provinent de les metadades pot resultar molt valuosa i, a més, el fet de poder verificar, a partir de la informació obtinguda, la seva (o no) autenticitat i integritat pot resultar de gran utilitat a l'hora de combatre *fake news*.

Com en el cas anterior, existeixen una gran varietat de recursos **OSINT** disponibles que permeten explotar les metadades d'imatges i vídeos, entre els quals trobem les web *Metadata2go*[44] i *theXifer*[45]. Altres recursos únicament proporcionen les metadades d'imatges però resulten interessants per l'estudi ja que, juntament amb aquestes metadades, faciliten la localització exacta en un mapa del lloc on es va efectuar la fotografia. Aquestes web són *Jimpl*[46], *Forensically*[47] i *Verexif*[48]. A més, *Verexif* ofereix la possibilitat d'eliminar o modificar les metadades de la imatge.

[44] https://www.metadata2go.com/view-metadata
[45] https://www.thexifer.net/
[46] https://jimpl.com/
[47] https://29a.ch/photo-forensics/#exif-meta-data
[48] https://www.verexif.com/en/

Figura 2. Metadades d'un vídeo obtingudes amb *theXfier*.

3.1.1.C. Geolocalització d'imatges i vídeos

Altres recursos **OSINT** poden ubicar geogràficament el lloc on es va enregistrar una fotografia o un vídeo. Això permet verificar si la ubicació es correspon amb la que es pretén en la publicació o notícia que les acompanya. Tanmateix, cal esmentar que el si dispositiu amb el qual es capta la imatge o vídeo té en la seva configuració la ubicació GPS desactivada en pot impossibilitar la seva geolocalització. Com en els casos anteriors, existeixen diversos recursos **OSINT** que permeten geolocalitzar imatges i/o vídeos. Començant pel mateix buscador *Google Maps*[49] el qual, a partir de les coordenades obtingudes en les metadades, pot establir la localització exacta. També existeixen altres pàgines web que, a partir de les mateixes imatges o vídeos, permeten ubicar-ne la posició. Exemples els trobem, a part dels esmentat en l'apartat anterior, en les web *Metadato*[50], que permet geolocalitzar imatges (i també explotar-ne les metadades), i *YouTube Geofind*[51], que permet geolocalitzar vídeos de *YouTube*. Un cop es té geolocalitzada la

[49] https://www.google.com/maps
[50] https://www.metadato.org/
[51] https://mattw.io/youtube-geofind/

imatge o el vídeo, es pot emprar l'eina *Google Street View*[52] per poder verificar visualment si realment el lloc es correspon al de la imatge o vídeo.

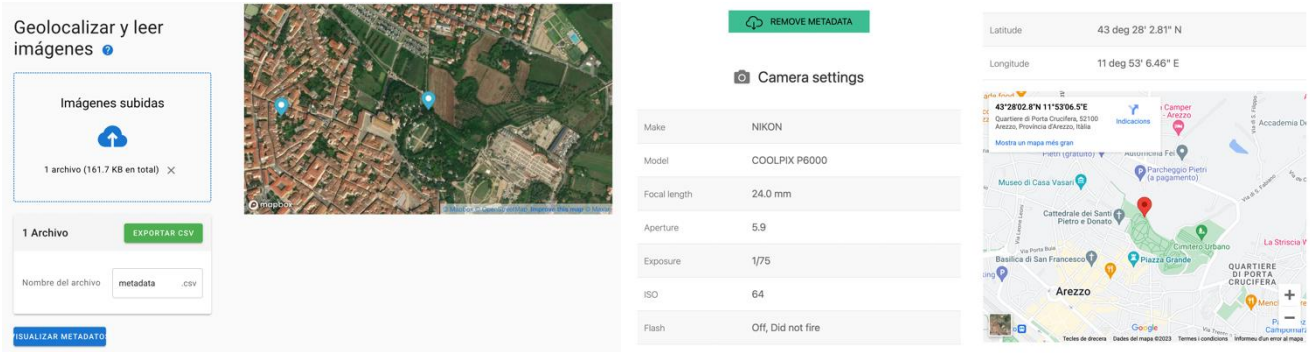


Figura 3. Geolocalització d'una imatge emprant *Metadato* i *Jimpl*.

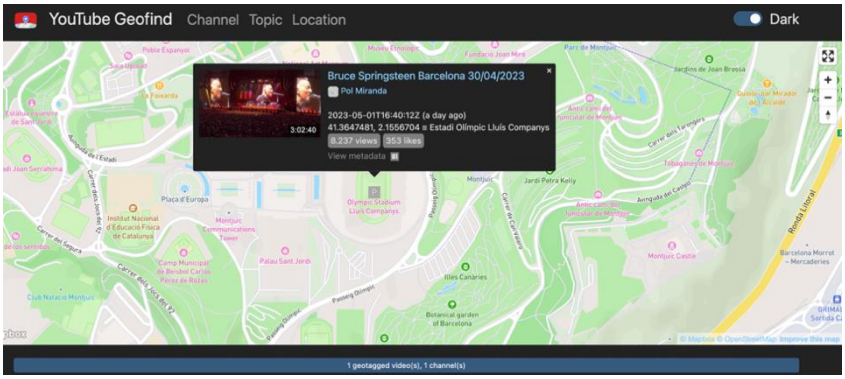


Figura 4. Geolocalització d'un vídeo de *YouTube* mitjançant la web Yo*uTube Geofind*.

3.1.1.D. Traducció d'imatges

Una altre recurs que pot permetre detectar *fake news* és la traducció d'imatges, ja que traduir una imatge que acompanya una notícia o una publicació en una xarxa social pot ajudar a contrastar si efectivament diu el que anuncia la informació que l'acompanya. Per dur a terme aquesta tasca, es poden fer servir els recursos que ofereixen *Google Translate*[53] i *Yandex*[54].

Figura 5. Traducció d'una imatge emprant *Google Translate*.

[52] https://www.google.com/streetview/
[53] https://translate.google.com/?sl=ru&tl=ca&op=images
[54] https://translate.yandex.com/

3.1.2. Anàlisi de perfils a les xarxes socials

Analitzar perfils en una xarxa social mitjançant recursos **OSINT** ajuda a reconèixer quins perfils o comptes poden ser susceptibles de ser un bot, és a dir; programes controlats automàticament que simulen el comportament humà en les xarxes socials interactuant amb altres perfils o comptes (Santana & Huerta Cánepa, 2019). Aquesta tasca es du a terme per mitjà de l'anàlisi de l'activitat dels comptes a la xarxa social, així com la identificació de patrons que poden indicar si les interaccions realitzades són constitutives d'un comportament humà o d'un robot automatitzat. En el cas de constatar que un compte o perfil en una xarxa social és un bot, augmenten molt les probabilitats que la informació que publica sigui *fake news*. Si ens centrem, a tall de mostra, en la xarxa social *Twitter*, existeixen diversos recursos **OSINT** que permeten analitzar i monitoritzar el comportament d'un compte així com, per exemple, els comptes que segueix, els que el segueixen, el nombre de tweets publicats, el tipus, el llenguatge emprat, els hashtags emprats, els retweets, les respostes i les citacions.

Entre la gran diversitat de recursos **OSINT** que permeten realitzar aquestes tasques, ens trobem amb les pàgines web *Botometer*[55], *Accountanalysis*[56] i *Tinfoleak*[57]. Cadascuna d'aquestes pàgines facilita tot tipus d'informació i, fins i tot, *Botometer* realitza una classificació en la que, a partir de l'activitat del compte, li dona una puntuació de l'1 al 5, on les puntuacions més altes són sinònim d'una activitat més semblant a la que realitza un bot.

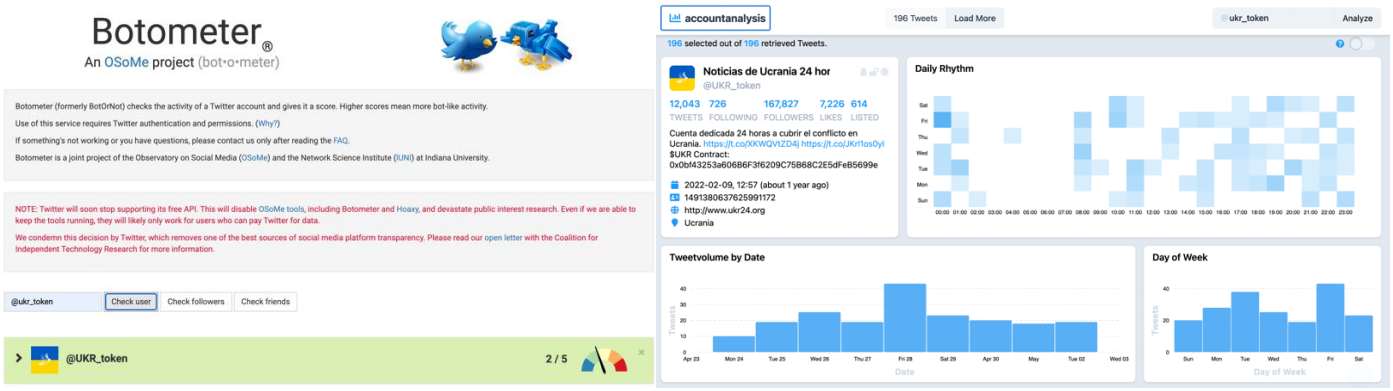


Figura 6. Anàlisi d'un perfil de *Twitter* emprant Bot*ometer* i *Accountanalysis*.

[55] https://botometer.osome.iu.edu/
[56] https://accountanalysis.app/
[57] https://tinfoleak.com/

3.1.3. InVID-WeVerify, el plugin OSINT per lluitar contra les *fake news*

Un cop detallats els diversos recursos **OSINT** disponibles online per qualsevol usuari que vulgui combatre les *fake news* escampades pel ciberespai, cal esmentar l'eina *InVID-WeVerify*[58], una extensió (també coneguda com a plugin) multi idioma desenvolupada pels navegadors web *Chrome* i *Firefox*. Aquesta eina està específicament dissenyada per ajudar als verificadors de fets a verificar els continguts publicats en el ciberespai, ajudant-los a estalviar temps i a ser més eficients en aquelles tasques enfocades a lluitar contra la desinformació i les *fake news*. En aquest sentit, segons el Poynter Institute[59], seu de la International Fact-Checking Network (**IFCN**), el connector de verificació InVID-WeVerify "*esdevé una de les eines més potents per detectar informació errònia en línia*"[60]. En aquest sentit, les eines de verificació disponibles en aquest plugin han estat mostrades en diversos tallers, inclòs en un organitzat per l'Observatori Europeu de Mitjans Digitals (**EDMO**) i, el setembre del 2021, va obtenir el primer premi de l'US Paris Tech atorgat pel Global Engagement Center del Departament d'Estat dels Estats Units d'Amèrica i el Digital Forensic Lab del Consell Atlàntic en una competició mundial entre quaranta eines de verificació.

Si s'analitzen les eines disponibles, el **mòdul de vídeo** permet analitzar vídeos de *YouTube*, *Facebook* o *Twitter*; fragmentar un vídeo de *YouTube*, *Facebook* o *Twitter* o un arxiu MP4 en fotogrames per realitzar una recerca inversa d'imatges a través de *Google*, *Yandex*, *Bing*, *Tineye*, *Baidu* o *Karma Decay*; extreure i analitzar els thumbnails, un miniatura de la imatge de mida original que s'utilitza quan és massa gran per mostrar-se com a vista prèvia d'un vídeo de *YouTube*; proporcionar informació sobre els drets legals d'un vídeo publicat a *YouTube* o *Twitter*; i extreure les metadades per imatges (en format JPEG) i vídeos (en format MP4 o M4V).

Seguidament, el **mòdul d'imatge** ofereix l'opció de proporcionar informació i analitzar el context d'una imatge publicada a les xarxes socials *Facebook* o *Twitter*; examinar a fons una imatge mitjançant una aplicació en forma de "lent d'augment" i un editor; extreure les metadades per imatges (en format JPEG) i vídeos (en format MP4 o

[58] https://chrome.google.com/webstore/detail/fake-news-debunker-by-inv/mhccpoafgdgbhnjfhkcmgknndkeenfhe?hl=en
[59] https://www.poynter.org/
[60] https://www.poynter.org/fact-checking/2021/these-6-tips-will-help-you-spot-misinformation-online/

M4V); detectar falsificacions i alteracions en imatges manipulades; llegir i traduir el text contingut en una imatge; i una funció avançada restringida a verificadors de fets, periodistes i investigadors registrats, per la qual és necessari registrar-se com a tals, que permet crear un GIF, un arxiu d'imatge animat que parteix de la combinació de diverses imatges, entre una imatge manipulada i la original per mostrar com ha estat manipulada.

Per la seva banda, el **mòdul de cerca** permet realitzar consultes de cerca avançada a la xarxa social *Twitter*; buscar *Factchecks*, és a dir, verificacions de fets ja realitzades; i realitzar cerques personalitzades de consultes entre xarxes en la que, a partir d'un text de cerca, es localitzen les dades associades a aquest a les xarxes i plataformes *Twitter*, *TikTok*, *8kun*, *YouTube*, *Linkedin*, *Reddit*, *Facebook*, VK, *Instagram* i *4chan*.

Finalment, el **mòdul d'anàlisi de dades** proporciona una funció avançada restringida a verificadors de fets, periodistes i investigadors registrats que proporciona una eina d'anàlisi de xarxes socials a *Twitter*; i, finalment, permet realitzar un anàlisi CSV de les xarxes socials *Facebook* i *Instagram* des d'una exportació de *CrowdTangle*, una eina d'estadístiques públiques que serveix perquè editors, periodistes, investigadors i verificadors de fets, entre altres, puguin registrar, analitzar i informar el que succeeix als mitjans socials. Entre altres funcionalitats, mostra les publicacions realitzades, permet saber amb quina freqüència es comparteix un enllaç en una xarxa social, qui el comparteix i els comentaris associats a aquest. *CrowdTangle* únicament registra dades de contingut públic i, per tant, no realitza cap seguiment de contingut que no sigui públic.

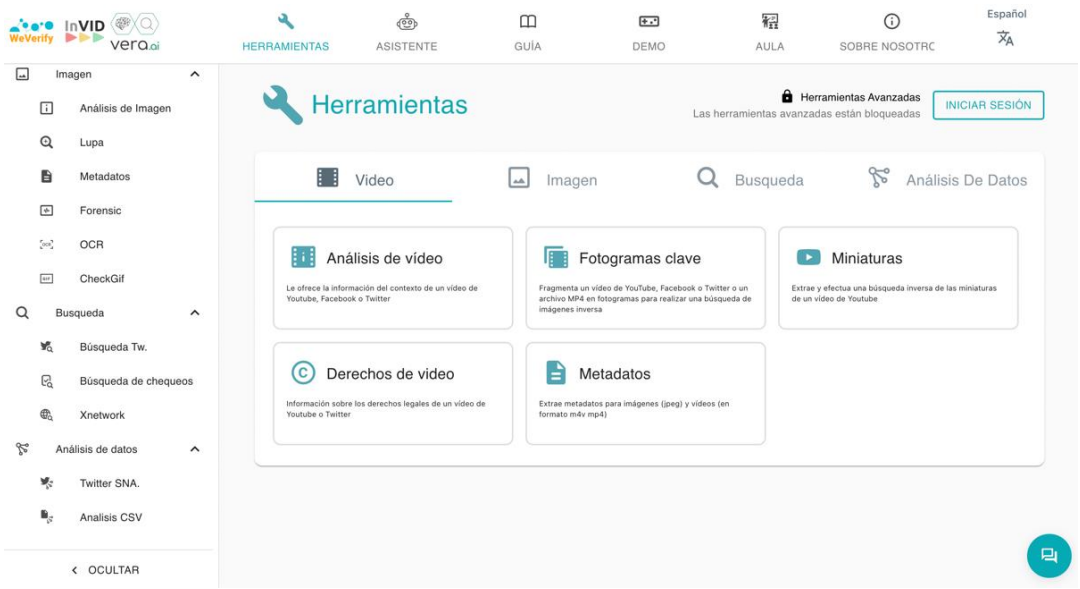


Figura 7. Vista de les eines disponibles al plugin *InVID-WeVerify*.

3.2. La figura dels fact-checkers o verificadors de fets

Els fact-checkers o verificadors de fets esdevenen un actor fonamental que juga un paper cabdal, juntament amb investigadors i altres actors civils, en la lluita contra la desinformació i les *fake news*, ja que funcionen com un mecanisme de resposta contra la seva propagació, sobretot quan aquesta es du a terme emprant com a mitjà de transmissió les noves tecnologies, Internet i les TIC. És per això diverses entitats han estat treballant per crear una xarxa d'alerta ràpida formada per verificadors de fets que sigui capaç d'informar de forma neutral, ràpida, eficaç i eficient a qualsevol persona que, navegant pel ciberespai, es trobi davant d'una *fake news*.

El fact-checking és una branca del periodisme que va aparèixer com un mecanisme de control intern i consisteix, bàsicament, en "*aplicar tècniques de periodisme per desemmascarar errors, ambigüitats, mentides, manca de rigor o inexactituds de continguts que han estat publicats als mitjans de comunicació*" (Ufarte-Ruiz, Peralta-García, & Murcia-Verdú, 2018). A més, amb l'expansió de les *fake news* relacionada, sobretot, amb la seva propagació pel ciberespai, s'ha introduït un element desestabilitzador i distorsionador que cal ser respost amb la màxima promptitud i eficàcia. Per això s'han desenvolupat sistemes de verificació de fets que, emprant eines tecnològiques i periodístiques, realitzen pràctiques per combatre la propagació de *fake news*, sobretot en el ciberespai, i garantir una cobertura informativa correcta (Graves, Deciding what's true: The rise of political fact-checking in American journalism, 2016).

Els orígens de la verificació de fets tal i com s'entén actualment es troben a principis del segle XX a Amèrica, on s'identifiquen els primers fact-checkers no oficials en aquells periodistes que es van atrevir a desafiar afirmacions provinents de les grans empreses farmacèutiques (Amazeen, 2020). En aquest context, l'any 2003 apareix la primera plataforma nord-americana de verificació de fets, *FactCheck.org*[61]. Des d'aleshores, la tasca de verificació de fets s'ha expandit arreu del món, arribant-se a convertir en un "*rar exemple d'un moviment genuïnament transnacional que reuneix professionals de molts sistemes de mitjans i cultures periodístiques, així com de l'acadèmia i les esferes civil i política*" (Graves, 2016).

[61] https://www.factcheck.org/

Tanmateix, hi ha tres moments que ajuden a explicar el gran creixement viscut en la pràctica de verificació de fets els darrers anys: que la web de fact-checking Politifact[62] guanyés el premi Pulitzer l'any 2009; l'arribada al poder de Donald Trump als Estats Units l'any 2017 i de Jair Bolsonaro al Brasil l'any 2019, i les campanyes de desinformació associades a aquells processos electorals, en els quals diverses organitzacions, entitats governamentals i organismes oficials van començar a ser conscients per primer cop sobre la perillositat de la difusió de *fake news*; i, finalment, la propagació de *fake news* relacionada amb la pandèmia de la COVID-19, on es va obrir una nova dimensió en col·locar temàtiques com la salut i la ciència en el centre de la desinformació (Moreno-Gil & Salgado-de Dios, 2023).

És per això que la figura dels verificadors de fets ha tingut una gran acceptació tant en les esferes periodístiques i socials com en les polítiques. Tot i això, no han mancat crítiques provinents de sectors com, per exemple, certes àrees de la política nord-americana que els han titllat de partidistes i de tenir una orientació progressista i antirepublicana, cosa que, segons aquestes crítiques, porta als verificadors de fets a prendre posicions, malmetre la reputació de candidats i intentar crear tendències (López-Pan & Rodríguez-Rodríguez, 2019). Per aquest motiu, els darrers anys han aparegut diversos projectes i organitzacions que s'han encarregat de certificar que els verificadors de fets compleixen uns requisits d'ètica, imparcialitat i professionalitat i, alhora, els han ajudat a perfilar, millorar i professionalitzar les seves tasques i així com a prendre consciència de la necessitat d'emprar una visió global en la seva funció.

3.2.1. L'International Fact-checking Network (IFCN)

L'any 2015 sorgeix la Xarxa Internacional de Verificació de Fets (**IFCN**[63]) de la mà del Poynter Institute, una escola de periodisme i també una organització de recerca sense ànim de lucre ubicada a Sant Petersburg (Florida, EEUU). L'**IFCN** cerca aglutinar la comunitat de verificadors de fets d'arreu del món i els defensors de la informació real en la lluita global contra la desinformació i les *fake news*, així com per promoure la implantació i aplicació de bones pràctiques en aquest camp.

[62] https://www.politifact.com/
[63] https://www.poynter.org/ifcn/

L'**IFCN** actualment està conformat per un total cent sis plataformes de verificació de fets repartides per tot el món, quatre de les quals són espanyoles: *EFE Verifica*[64], *Maldita.es*[65], *Newtral*[66] i *Verificat*[67]. **L'IFCN** és un referent mundial en l'excel·lència de la verificació de fets a través de la promoció, formació i la realització d'esdeveniments globals. És per això que proporciona recursos per adoptar els estàndards bàsics i un codi de principis; impulsa subvencions i beques per ajudar als verificadors de fets i altres organitzacions a llençar programes i iniciatives noves i innovadores; realitza formacions en línia i presencials per ajudar als verificadors de fets a desenvolupar noves habilitats i capacitats; supervisa l'elaboració de polítiques globals sobre la verificació de fets i analitza les tendències així com els darrers esdeveniments ocorreguts mitjançant articles, butlletins i un informe anual que ofereixen recursos als verificadors de fets; i, finalment, connecta als verificadors de fets d'arreu del món mitjançant xerrades i esdeveniments com, per exemple, l'International Fact-Checking Day[68], que es celebra cada dos d'abril.

L'**IFCN** defensa que "*la veritat i la transparència poden ajudar la gent a estar millor informada i equipada per navegar per la desinformació perjudicial*" i per això, com ja s'ha esmentat, crea un codi de principis[69] format per cinc punts:

- El compromís pel no partidisme i l'equitat, que consisteix en emprar el mateix estàndard per cada verificació de fets, seguint sempre el mateix procés i deixant que les evidències siguin les que dictin les conclusions. Tampoc es defensaran ni adoptaran posicions polítiques sobre els temes que estiguin comprovant.
- El compromís amb els estàndards i la transparència de les fonts, que implica proporcionar totes les fonts emprades amb suficient detall perquè qualsevol usuari pugui, al seu torn, verificar i, en cas que ho cregui convenient, replicar el treball realitzat, amb l'única excepció que la seguretat d'alguna font es vegi compromesa. En aquest cas s'aportaran tants detalls com sigui possible sense comprometre la font.
- El compromís per la transparència en el finançament i la organització, a partir del qual es garanteix la transparència de les fonts de finançament, així com detallar la formació professional de les figures clau de la organització, la seva estructura

[64] https://verifica.efe.com/
[65] https://maldita.es/
[66] https://www.newtral.es/zona-verificacion/fact-check/
[67] https://www.verificat.cat/
[68] https://www.poynter.org/event/international-fact-checking-day/
[69] https://ifcncodeofprinciples.poynter.org/

organitzativa i el seu estatus jurídic. A més, també s'haurà d'indicar de forma clara un canal a partir del qual els lectors s'hi puguin comunicar.

- El compromís amb els estàndards i la transparència de la metodologia, pel qual s'ha de mostrar la metodologia emprada per seleccionar, investigar, escriure, editar i corregir les verificacions de fets. A més, s'anima als lectors a enviar reclamacions s'assegura la transparència sobre per què i com comproven els fets.
- El compromís amb una política de correccions oberta i honesta, que implica publicar i seguir escrupolosament la política de correccions. També es corregirà en tot moment i de manera clara i transparent d'acord amb la política de correccions, procurant, en la mesura del possible, garantir que es tingui accés a la versió corregida.

Per poder entrar a formar part de l'**IFCN** cal adherir-se a aquest codi de principis, demostrar que s'empra una metodologia sòlida i transparent, es treballa amb fonts fiables i no es tenen vincles amb partits polítics. Cada candidat, a més, haurà de superar una avaluació sobre el compliment del codi de principis i, en cas que sigui aprovada, cada any haurà de tornar a sol·licitar i enviar una carta de presentació per renovar la seva verificació superant, de nou, un procés d'avaluació. Formar part de **l'IFCN** implica estar compromès amb compliment dels fonaments deontològics que garanteixen la imparcialitat i professionalitat a l'hora de dur a terme la verificació de fets per combatre les *fake news*.

Un exemple de les iniciatives endegades per l'**IFCN** o algun dels seus membres el trobem amb el projecte *UkraineFacts*[70], creat a partir que la plataforma de verificació espanyola *Maldita.es*, després de constatar la gran quantitat de *fake news* i desinformació relacionades amb la invasió Russa a Ucraïna i la velocitat en la que es propagaven pel ciberespai, considerés necessària la creació d'una base de dades mundial que recollís de forma ràpida tota aquella desinformació. Aquesta iniciativa va rebre el suport d'altres membres de l'**IFCN** i, per això, es va decidir obrir l'accés a la base de dades a més de cent verificadors de fets d'arreu del món que en formen part. Com a resultat, el projecte facilita a qualsevol usuari que hi estigui interessat una pàgina web amb un mapa mundial interactiu que permet consultar les *fake news* i desinformació relacionades amb la invasió Russa a Ucraïna segons el país on s'hagin detectat, així com els desmentits realitzats per les diverses organitzacions de verificacions de fets que ho han investigat.

[70] https://ukrainefacts.org/

3.2.2. L'European Digital Media Observatory (EDMO)

L'Observatori Europeu de Mitjans Digitals (**EDMO**[71]), que va entrar en funcionament l'1 de juny del 2020, és un observatori independent finançat per la Comissió Europea que serveix de centre i aglutina una comunitat independent i multidisciplinària, formada per verificadors de fets i investigadors acadèmics amb experiència en el camp de la desinformació en línia que, en col·laboració amb mitjans de comunicació, plataformes en línia, professionals de l'alfabetització mediàtica i autoritats reguladores nacionals, treballa per combatre, a nivell europeu, la desinformació i les *fake news* al ciberespai.

L'**EDMO** té la seva base a l'Escola de Governació Transnacional de l'Institut Universitari Europeu a Florència (Itàlia) i compta amb una estructura de govern independent de qualsevol autoritat pública. El seu principal objectiu és coordinar projectes enfocats a identificar la desinformació i les *fake news* al ciberespai, desarrelar les seves fonts, diluir-ne l'impacte, donar suport a la verificació dels fets i la informació de qualitat, i interconnectar les diverses comunitats d'experts que en formen part, inclosa la comunitat de verificadors de fets europeus, entre els quals es troben els espanyols *Maldita*, *Newtral*, *VerificaRTVE*[72] i *Verificat*.

En un primer moment, l'**EDMO** va centrar la seva activitat en desplegar aquells mitjans tècnics i infraestructures necessaris per facilitar la col·laboració i l'intercanvi de coneixements entre investigadors i verificadors de fets a Europa, a més de coordinar aquelles activitats conjuntes realitzades a nivell europeu. Això inclou un programa de formació multidisciplinari i transversal enfocat, entre altres, a conèixer les tendències en l'àmbit de la desinformació i en potenciar les pràctiques de verificació de fets. Entre aquests programes en destaquen, a l'estar relacionades amb el present treball, la formació en línia dedicada a descobrir campanyes de desinformació a través de recursos **OSINT**[73], i la formació en línia dedicada a dominar l'ús del plugin *InVID-WeVerify*[74].

[71] https://edmo.eu/
[72] https://www.rtve.es/noticias/verificartve/
[73] https://edmo.eu/2022/12/01/edmo-online-training-uncovering-disinformation-campaigns-with-osint-tools/
[74] https://edmo.eu/2022/09/02/edmo-online-training-the-invid-weverify-verification-toolbox/

En una segona fase, es van crear un total vuit observatoris nacionals (on cada observatori pot abastar més d'un país), independents de qualsevol autoritat pública nacional o de la Unió Europea. Aquests observatoris van passar a formar part de la xarxa de l'**EDMO** amb l'objectiu d'augmentar la seva capacitat d'abordar campanyes de desinformació a nivell nacional i de la Unió Europea i a analitzar-ne el seu impacte. En aquest sentit, cada observatori ha de contribuir a la creació d'una comunitat multidisciplinària d'investigadors acadèmics, verificadors de fets, professionals dels mitjans de comunicació i altres parts interessades rellevants per tal de crear una xarxa capaç de detectar i analitzar campanyes de desinformació, organitzar activitats d'alfabetització mediàtica a nivell nacional o internacional i donar suport a les autoritats nacionals per al seguiment de les polítiques de les plataformes en línia i l'ecosistema dels mitjans digitals.

Un d'aquests observatoris és **IBERIFIER**[75], l'observatori de mitjans digitals d'Espanya i Portugal, coordinat des de la Universitat de Navarra. **IBERIFIER** està format per sis universitats, cinc organitzacions de verificació de fets i agencies de notícies i sis centres d'investigació multidisciplinari, i té com a missió principal analitzar l'ecosistema ibèric dels mitjans digitals i fer front al problema de la desinformació i les *fake news*. Entre moltes altres funcionalitats, **IBERIFIER** proporciona en la seva pàgina web un llistat[76] de totes les verificacions de fets realitzades pels equips de fact-checking que en formen part, els quals són el portuguès *Polígrafo*[77] i els espanyols *EFE Verifica*, *Verificat* i *Maldita.es*. També facilita informes trimestrals que mostren les narratives de desinformació detectades a Espanya i Portugal, sent el darrer de maig de 2023[78].

3.2.3. L'European Fact-checking Standards Network (EFCSN)

La Xarxa Europea d'Estàndards de Verificació de Fets (**EFCSN**[79]), impulsada per la Comissió Europea i presentada oficialment el 16 de novembre del 2022, és una associació formada per organitzacions de verificació de fets compromeses en mantenir i promoure els estàndards més alts en la seva tasca, i per crear vincles professionals i duradors entre la comunitat de verificadors de fets independents a nivell europeu. Per

[75] https://iberifier.eu/
[76] https://iberifier.eu/factchecks/
[77] https://poligrafo.sapo.pt/fact-checks
[78] https://iberifier.eu/2023/05/26/fact-checking-report-q1-2023/
[79] https://efcsn.com/

això, les organitzacions que conformen l'**EFCSN** es comprometen a complir amb els estàndards d'independència, transparència i qualitat periodística descrits al **Codi Europeu d'Estàndards per a Organitzacions Independents de Verificació de Fets**[80], un conjunt de criteris dissenyats per garantir que les organitzacions de verificació de fets es regeixen pels estàndards més alts de metodologia, ètica i transparència que les permeti, entre altres, liderar el combat contra la desinformació i les *fake news* a nivell europeu. És per això que si una organització vol esdevenir membre de l'**EFCSN**, haurà de verificar abans que compleix amb els estàndards establerts en aquest **Codi**, aprovat l'agost de l'any 2022 i que actualment es troba en fase d'adhesió i implantació.

L'**EFCSN** també vol integrar a la seva xarxa la comunitat europea d'organitzacions **OSINT**[81]**,** acadèmics i investigadors especialitzats en la verificació de fets, i clients de serveis de verificació de fets. En relació a les organitzacions **OSINT**, l'objectiu és que ajudin en el desenvolupament d'aplicacions i eines informàtiques específicament dissenyades per combatre la propagació de desinformació i *fake news* al ciberespai, així com crear i implementar programes de formació de qualitat que permetin als verificadors de fets, acadèmics i investigadors dominar l'ús dels recursos i tècniques **OSINT**. Per això l'**EFCSN** estableix un objectiu comú dels verificadors de fets i la comunitat **OSINT** per al desenvolupament d'aquelles pautes estandarditzades que guiïn i ajudin a dur a terme la seva tasca i missió.

Actualment, **l'EFCSN** està formada per un total de vint-i-nou organitzacions de verificació de fets d'arreu d'Europa, entre les quals hi ha les espanyoles *Maldita*, que també forma part del seu Consistori de govern, *Newtral*, *EFE Verifica* i *Verificat*. Tanmateix, de l'1 al 31 de maig d'aquest any es va obrir una segona ronda per incorporar-ne de noves i es preveu una nova ronda per l'octubre d'aquest any.

[80] https://eufactcheckingproject.com/app/uploads/2022/10/EU-CODE-EFCSN-.pdf
[81] https://eufactcheckingproject.com/get-involved-osint-orgs/

4. DISCUSSIÓ I CONCLUSIONS

Les *fake news*, tot i que no són, ni molt menys, un fenomen nou, han esdevingut els darrers temps un dels perills més reals i importants al qual les societats i governs democràtics d'arreu del món s'han hagut d'enfrontar. Això és degut, entre altres factors, a que les *fake news* han sabut aprofitar el nou medi que el ciberespai les ha ofert per expandir-se de forma exponencial, cosa que les ha permès arribar a un nombre potencial de persones que a través dels mitjans convencionals no hagués estat possible. A aquest fet cal afegir que, cada cop més s'utilitza el ciberespai en general, i les grans plataformes d'Internet i xarxes socials en particular, com a font d'informació, cosa que ha propiciat que, aprofitant aquesta nova conjuntura, les *fake news* hagin estat l'eina emprada per condicionar, manipular i, fins i tot, alterar percepcions i dirigir opinions cap a una direcció deliberadament buscada per aquells que estan darrera de la seva creació i divulgació.

Per aquests motius, la lluita contra la ciberamenaça que representen les *fake news* ha esdevingut una de les grans prioritats per organitzacions, administracions, autoritats i governs, tant a nivell nacional com internacional. Per dur a terme aquesta tasca, s'han endegat una pluralitat d'iniciatives per tal de poder combatre-hi de forma efectiva, tant a nivell legal, on es busca exercir un control formal amb la inclusió o adaptació de preceptes legals que recullin les conductes realitzades a l'hora de crear i difondre *fake news* i així trobar-hi un encaix legal, com a nivell pràctic, on la tecnologia juga un paper crucial assumint el rol de control informal a través de la utilització d'una gran varietat d'eines i aplicacions informàtiques que permeten localitzar, identificar, desemmascarar i rebatre les *fake news*, així com la creació i desenvolupament de noves específicament dissenyades per dur a terme aquesta tasca.

En l'àmbit legal, s'ha fet un estudi descriptiu de la legislació espanyola i la europea on s'han pogut constatar les diferències d'enfocament existents entre ambdues, així com possibles mancances detectades a nivell espanyol en relació al tractament penal que es dona a les *fake news*. Tot i això, cal destacar que la conjunció entre ambdues legislacions pot esdevenir, un cop entri completament en vigor la **DSA**, una eina que pot ser realment eficaç per combatre les *fake news* al ciberespai.

Amb això, la legislació espanyola ha optat per perseguir i castigar mitjançant l'aplicació del dret penal aquelles conductes individuals que una persona (física i, de forma més restringida, jurídica) pot arribar a cometre quan crea, publica i difon *fake news*. En aquest sentit, a diferència de l'enfocament europeu, on s'han desenvolupat noves lleis específicament pensades per combatre les campanyes de desinformació i les *fake news* al ciberespai, en el dret penal espanyol no existeix una tipificació expressa que contempli la creació i divulgació de *fake news* i, per tant, aquestes actituds només resulten punibles quan mitjançant la creació i divulgació de *fake news* es cometen delictes que si es troben tipificats en el **Codi Penal**. Aquest fet pot portar a possibles confusions en tant es podria considerar que els preceptes legals emprats per perseguir i castigar la creació i difusió de *fake news* esdevenen, en ocasions, gairebé impossibles d'aplicar a la pràctica. Com a mostra estan els delictes de desordres públics, en tant que resulta altament improbable que la difusió d'una *fake news* pugui generar una situació en la que sigui necessari prestar auxili o provoqui una mobilització efectiva dels serveis d'emergència.

A això cal afegir-hi la dificultat d'abordar temes tan importants com la redifusió de *fake news* al ciberespai, en tant que aquest esdevé un mitjà on poden propagar-se de manera gairebé il·limitada, cosa que afavoreix la seva potencial viralitat. En aquest sentit, a diferència del que succeeix amb el segon paràgraf de l'**article 197.7 CP**, on es castiguen les conductes relacionades amb el reenviament de *sexting* sense consentiment de la víctima, qui realitza aquesta redifusió pot no ser conscient que el que està retransmetent és, precisament, una *fake news* en la que, a més, sense conèixer la intenció amb la que aquesta va ser inicialment divulgada, pot estar contribuint, de forma involuntària, a la creació d'un determinat resultat delictiu. És per això que es poden plantejar seriosos problemes d'imputació subjectiva a l'hora d'intentar perseguir aquesta redifusió.

D'altra banda, a nivell europeu s'ha optat per enfocar aquesta lluita des d'una perspectiva més global, en el sentit que no es busca castigar fets comesos de forma individual, sinó que l'objectiu és combatre possibles ingerències d'organitzacions i/o països tercers que, a través de les *fake news* i les campanyes de desinformació, busquen desestabilitzar a gran escala el normal funcionament de les institucions i governs que formen part de la Unió Europea. A més, cal recordar que el dret penal esdevé patrimoni dels estats i, per tant, a nivell europeu es requereix un enfocament legal des d'una perspectiva diferent. És per això que, en aquest cas, el legislador ha posat el focus en els

grans prestadors de serveis digitals, en tant que ha entès que aquests són l'eix vertebrador i indispensable que proporciona a les *fake news* un entorn on poder-se propagar, gairebé, de forma il·limitada i incontrolada. En aquest context, mitjançant els diversos plans i comissions creats i desenvolupats, però, sobretot, el **Codi de bones pràctiques en matèria de desinformació** i la **DSA**, es cerca garantir que les grans plataformes i proveïdors d'Internet empren totes aquelles les mesures de les que disposen per lluitar contra els riscs causats per les *fake news* i les campanyes de desinformació que utilitzen els serveis que ofereixen com a mitjà de propagació. Aquest és, doncs, un enfocament que pot esdevenir certament útil en tant que ajuda a limitar la capacitat de propagació de les *fake news* i les campanyes de desinformació al ciberespai.

Per assolir aquests objectius, la **DSA** estableix un nou marc normatiu aplicable als prestadors de serveis i intermediaris en línia en el qual es promou un servei més segur i equitatiu que ajuda a combatre, mitjançant l'establiment de les corresponents responsabilitats i obligacions, la divulgació de continguts il·lícits i/o campanyes de desinformació que puguin esdevenir un risc per la societat, entre els que s'inclouen aquells efectes negatius que puguin incidir en la salut i/o la seguretat públiques, el discurs civil, la participació política i la igualtat. Per la seva banda, tot i que l'adhesió al **Codi** és voluntària, amb l'entrada en vigor de la **DSA** esdevé la forma més senzilles per demostrar que les grans plataformes i proveïdors de serveis compleixen amb els requisits establerts en aquesta. És per això que la **DSA** i el **Codi** esdevenen una veritable punta de llança per combatre la creació i divulgació de *fake news* al ciberespai.

A simple vista, doncs, es pot apreciar la diferència d'enfocament entre la legislació espanyola i la europea, en tant que l'espanyola es centra a castigar actituds i accions individuals, mentre que l'europea està centrada en evitar que el ciberespai esdevingui un camp propici per la divulgació de *fake news*. En aquest sentit, mentre a nivell espanyol la pena s'imposa a la persona (física o jurídica) responsable de la creació i divulgació d'una *fake news*, a nivell europeu es castiga amb fortes penes de multa, que poden arribar al 6% de la seva facturació mundial, aquells proveïdors de serveis digitals que no disposin dels medis suficients, eficients i necessaris per combatre la propagació de *fake news* i campanyes de desinformació que utilitzen els serveis que ofereixen al ciberespai com a mitjà de transmissió.

Amb tot, es pot afirmar que la conjunció entre la normativa espanyola i l'europea, ja que la **DSA**, en ser una norma europea amb rang de Reglament, és vinculant per tots els països que formen part de la Unió Europea, esdevé una eina molt complerta que ha de permetre combatre la creació i propagació de *fake news* al ciberespai. Tot i això, s'entén necessari endegar una reforma del **Codi Penal** que permeti cobrir aquelles llacunes que actualment poden dificultar la lluita contra les *fake news*. En aquest sentit, tipificar nous delictes de *fake news* i/o introduir agreujants a delictes ja existents serien algunes de les vies que podrien ajudar a cobrir les mancances actuals. D'altra banda, un altre factor a tenir en compte és el temps que pot arribar a trigar l'estat espanyol en transposar i adaptar a l'ordenament jurídic espanyol la **DSA**. Per aquest motiu, al qual s'ha d'afegir la proximitat temporal des que va ser aprovada, serà necessari esperar per poder apreciar la verdadera efectivitat de la **DSA**, cosa que podrà ser tractada en futures investigacions.

Pel que fa al segon gran eix vertebrador en la lluita contra les fake news, la vessant tecnològica, aquest treball ha mostrat, per un costat, recursos i tècniques **OSINT** emprats per combatre les *fake news* i, per l'altra, ha estudiat la figura dels **verificadors de fets** que, tot i no ser per si mateixos una eina tecnològica, si que, entre altres tècniques, fan ús de la tecnologia al seu abast per poder realitzar la seva tasca de forma eficaç.

En relació als recursos i tècniques **OSINT** per lluitar contra les *fake news*, s'han treballat aquelles a les que qualsevol usuari pot tenir al seu abast quan navega pel ciberespai i que, alhora, resulten senzilles d'emprar, en el sentit que amb una mica de pràctica o realitzant una formació bàsica ja poden resultar realment útils. Partint d'aquesta base, s'han mostrat aquelles eines que poden resultar més efectives i que, alhora, són les més emprades per combatre les *fake news*. Entre les estudiades destaca el plugin inVID-WeVerify, una eina de verificació de continguts específicament dissenyada per combatre les *fake news* que, al seu torn, incorpora una gran varietat d'eines **OSINT**.

Aquest plugin és emprat, entre altres, per **verificadors de fets**, una figura que esdevé gairebé fonamental per lluitar contra les *fake news*, ja que el seu objectiu primordial és cercar, desemmascarar i contrarestar les *fake news* divulgades, en la gran majoria d'ocasions, pel ciberespai. A més, poden realitzar aquesta tasca de forma gairebé immediata, ja que els equips que els conformen tenen la capacitat de detectar l'aparició d'una possible *fake news* i, a partir d'aquest moment, endegar el procés de verificació que

permetrà investigar i contrastar la informació publicada i, si és necessari, desmentir-la. Com a exemple tenim el projecte *UkraineFacts*, nascut per combatre ràpida i eficaçment aquelles *fake news* divulgades al ciberespai relacionades amb la invasió Russa a Ucraïna.

Una mostra de la rellevància que tenen els **verificadors de fets** es pot apreciar en el fet que han estat inclosos en textos, tractats i legislació europeus, entre els que destaquen el **Codi de bones pràctiques de la Unió en matèria de desinformació** i la **DSA** que, per una banda, busquen potenciar i facilitar la seva tasca i, per l'altra, garantir la seva neutralitat, independència i transparència. Aquest és un dels motius pels quals han anat sorgint institucions internacionals com **IFCN**, l'**EDMO** o l'**EFCSN**, les quals, a partir de l'adhesió i compliment dels seus codis de principis i conductes, garanteixen que els **verificadors de fets** que en formen part es regeixen pels estàndards més alts de metodologia, ètica i transparència.

En aquest punt, s'ha de destacar la gran quantitat i diversitat de recursos que, des de múltiples i variats àmbits, s'estan abocant a la lluita contra les *fake news*. Tots aquests esforços són una bona evidència de la importància que els governs, administracions, organitzacions i institucions, públiques i privades, han donat al poder d'influència que, tant a nivell nacional com internacional, les *fake news* poden arribar a tenir, així com el perill potencial que representen per la societat.

Tanmateix, cal preguntar-se si l'esforç realitzat està sent suficient. En aquest sentit, cal dir que, tot i que resulta impossible erradicar les *fake news* de forma complerta, aquest treball ha demostrat que es disposa d'unes eines legals i tecnològiques amb el potencial suficient per combatre-les d'una forma realment eficaç. Així doncs, per una banda, els recursos tecnològics i, sobretot, els **verificadors de fets**, permeten donar una resposta ràpida en el moment en què es detecta l'aparició d'una nova *fake news* i, de l'altra, la **DSA** a nivell europeu i el **Codi Penal** a nivell espanyol, estan per garantir que el ciberespai no esdevingui un lloc en el qual es pugui difondre *fake news* de forma impune i incontrolada.

Tot i això, s'ha de tenir en compte que aquesta batalla, com tot el que envolta el ciberespai i la seva evolució tecnològica permanent, és una batalla dinàmica, ja que, de ben segur, els creadors de *fake news* trobaran noves formes i mitjans per generar-les,

propagar-les i amagar la seva identitat. En aquest context, una tecnologia que cal tenir molt en compte és la intel·ligència artificial que, mitjançant eines com **ChatGPT**, de forma inexorable comencen a obrir-se pas i a ser més accessibles per tot tipus d'usuaris. Aquesta tecnologia pot esdevenir una amenaça pels riscs que pot comportar-ne el seu mal ús a l'hora de crear i propagar *fake news*, cosa que ja està provocant els primers moviments parlamentaris per regular-ne el seu desenvolupament i ús. Tot i això, alhora, també pot ser, si se sap utilitzar de forma correcta, una eina molt potent per combatre, precisament, les *fake news*. Per aquests motius, serà interessant seguir-ne la seva evolució en un futur proper.

Finalment, no voldria acabar aquest treball sense esmentar una frase que, com a bon aficionat a les sèries, em va impactar en el moment de sentir-la i que, a més, va ser una de les llavors que va fer que em decantés a realitzar un treball que tractés en la seva temàtica les *fake news*. Aquesta frase, extreta de la mini sèrie **Chernobyl**, diu: "*Quin és el cost de les mentides? No és que els confondrem amb la veritat. El veritable perill és que si escoltem prou mentides, ja no reconeixerem la veritat en absolut*".

BIBLIOGRAFIA

Amazeen, M. A. (2020). Journalistic interventions: The structural factors affecting the global emergence of fact-checking. *Journalism Vol. 21(1)*, 95–111.

Fernández Entralgo, J. (2020). Injúrias y calumnias. A N. Rodríguez Gutierrez, A. Campomanes Caleza, J. Fernández Entralgo, J. López Ordiales, P. Mora Díez, R. Rodríguez Ruiz, & M. E. San José Asensio, *Análisis jurídico de las fake news en los tipos penales* (p. 107-124). Madrid: Editorial Jurídica Sepín S.L.

Fernández Fernández, M. (2020). Informaciones, crónicas y relaciones del siglo XVI como antecedentes de las fake news. El caso de la "invención" de San Segundo. *Historia y comunicación social, 26(2)*, 593-602.

Fernández-García, N. (2017). Fake news: una oportunidad para la alfabetización mediática. *Nueva sociedad, (269)*.

Fuentes-Lara, C., & Arcila-Calderón, C. (2023). El discurso de odio islamófobo en las redes sociales. Un análisis de las actitudes ante la islamofobia en Twitter. *Revista Mediterránea de Comunicación*.

Gelfert, A. (2018). Fake news: A definition. *Informal logic, 38(1)*, 84-117.

Graves, L. (2016). Deciding what's true: The rise of political fact-checking in American journalism. *Columbia University Press*.

Graves, L. (2016). Mapping the institutional roots of the global fact-checking movement. *Journalism studies, 19(5)*, 613-631.

Gómez, L. (2011). Un espacio para la investigación documental. *Revista Vanguardia psicológica clínica teórica y práctica*, 226-233.

Llinares, F. M. (2012). A F. M. Llinares, *El cibercrimen: Fenomenología y criminología de la delincuencia en el ciberespacio.* Marcial Pons.

López-Pan, F., & Rodríguez-Rodríguez, J. M. (2019). El Fact Checking en España. Plataformas, prácticas y rasgos distintivos. *Estudios sobre el Mensaje Periodístico*.

Mora Díez, P. (2020). La difusión indiscriminada y masiva de intimidad de las personas a través de las fake news y su relevancia penal. A N. Rodríguez Gutierrez, A. Campomanes Caleza, J. Fernández Entralgo, J. López Ordiales, P. Mora Díez, R. Rodríguez Ruiz, & M. E. San José Asensio, *Análisis jurídico de las fake news en los tipos penales* (p. 51-65). Madrid: Editorial Jurídica Sepín S.L.

Morejón-Llamas, N., Martín-Ramallal, P., & Micaletto-Belda, J.-P. (2022). Curación de contenido en Twitter como antídoto a la guerra híbrida durante la invasión de Rusia a Ucrania. *Profesional de la información, vol, 31, núm. 3*.

Moreno-Gil, V., & Salgado-de Dios, F. (2023). El cumplimiento del código de principios de la International Fact-Checking Network en las plataformas de verificación españolas. Un análisis cualitativo. *Revista de Comunicación, 22(1)*, 293-307.

Odar, R. M. (2016). Tipología de las investigaciones jurídicas. *Derecho y cambio social, 13 (43)*, 10.

Palacios, R. M. (2006). Investigación cualitativa y cuantitativa: Diferencias y limitaciones. Puira, Perú.

Parra Valero, P., & Oliveira, L. (2018). Fake news: una revisión sistemática de la literatura (Fake news: a systematic review of the literatura). *Observatorio (OBS*), 12(5)*, 54-78.

Parrondo, L. (2017). Tecnología blockchain, una nueva era para la empresa.

Quinto Huamán, C., Armas Vega, E. A., Sandoval Orozco, A. L., & García Villalba, L. J. (2016). Análisis de metadatos en vídeos digitales de dispositivos móviles.

Rodríguez Gutiérrez, N. (2020). Delitos contra el mercado y los consumidores. A N. Rodríguez Gutierrez, A. Campomanes Caleza, J. Fernández Entralgo, J. López Ordiales, P. Mora Díez, R. Rodríguez Ruiz, & M. E. San José Asensio, *Análisis jurídico de las fake news en los tipos penales* (p. 149-161). Madrid: Editorial Jurídica Sepín S.L.

Rodríguez Pérez, C. (2019). No diga fake news, di desinformación: una revisión sobre el fenómeno de las noticias falsas y sus implicaciones. *Comunicación (40)*, 65-74.

Salas Abad, C. (2019). La primera fake news de la historia. *Historia y comunicación social, 24(2)*, 411.

Salaverría, R., Buslón, N., López-Pan, F., León, B., López-Goñi, I., & Erviti, M.-C. (2020). Desinformación en tiempos de pandemia: tipología de los bulos sobre la Covid-19. *El profesional de la información (EPI), 29(3)*.

Santana, L. E., & Huerta Cánepa, G. (2019). ¿Son bots? Automatización en redes sociales durante las elecciones presidenciales de Chile 2017. *Cuadernos. info, (44)*, 61-77.

Scott, J. (2014). Evidence and data in social research. En *A matter of record: Documentary sources in social research.* John Wiley & Sons.

Sintes-Olivella, M., Xicoy-Comas, E., & Yeste-Piquer, E. (2020). Blockchain al servicio del periodismo de calidad. El caso Civil. *Profesional de la información, v. 29, n. 5*, 28.

Sun-Tzu. (2016). *El arte de la guerra.* Aegitas.

Tapscott, D., & Tapscott, A. (2018). *Blockchain revolution.* Senai-SP Editora.

Toro-Alvarez, M. M., Jaimes, W. D., & Bonilla-Duitama, M. L. (2018). Investigación del Cibercrimen y de los Delitos Informáticos Utilizando Inteligencia de Fuentes Abiertas de Información (OSINT).

Ufarte-Ruiz, M.-J., Peralta-García, L., & Murcia-Verdú, F.-J. (2018). Fact checking: un nuevo desafío del periodismo.

Valcárcel Siso, R. L., Carrascal Domínguez, S., Pintado, A., & Nicolás, M. J. (2020). La Unión Europea ante la desinformación y las fake news. El fact checking como un recurso de detección, prevención y análisis. A *Aproximación periodística y educomunicativa al fenómeno de las redes sociales* (p. 985-1002). McGraw-Hill Interamericana de España.

Wardle, C., & Derakhshan, H. (2018). Thinking about 'information disorder': formats of misinformation, disinformation, and mal-information. *Journalism, 'fake news' & disinformation*, 43-54.

ANNEX A. CASOS REALS DE CONDEMNES PER DELICTES D'ODI PER LA DIFUSIÓ DE *FAKE NEWS* AL CIBERESPAI

A l'informe de la Secretaria Tècnica de la FGE sobre el tractament penal que han de tenir les *fake news*, s'esmenten dues querelles presentades per la Fiscalia Provincial de Barcelona contra dos usuaris de la xarxa social *Twitter* per haver publicat missatges a l'esmentada xarxa social atribuint falsament fets delictius a un determinat col·lectiu, els MENAS o menors estrangers no acompanyats, format per menors de 18 anys, immigrants que en la seva gran majoria han arribat des del nord d'Àfrica, han entrat al país de forma irregular i que no tenen cap familiar directe que se'n pugui fer càrrec. En aquest sentit, l'informe de la Secretaria Tècnica de la FGE destaca la voluntat de les dues persones querellades de menyscabar i generar descrèdit a aquest col·lectiu en concret.

ANNEX A.1. Primera querella a Espanya per un delicte d'incitació a l'odi per difondre una *fake news* emprant les xarxes socials

Aquest primer cas s'inicia quan el dia dos de juliol de 2019 la Síndica de Greuges de Barcelona va presentar una denúncia davant la Fiscalia Provincial de Barcelona per la difusió a través d'Internet i les xarxes socials, concretament la xarxa social *Twitter*, d'un vídeo amb expressions que inicialment podrien encaixar en l'**article 510.2.b) CP**, a l'associar de forma esbiaixada i interessada al col·lectiu de MENAS amb la violència als centres escolars.

Concretament, la persona usuària de la xarxa social *Twitter* va publicar al seu compte un tuit amb un vídeo on es veia com a una aula d'un centre educatiu, diversos menors cridaven i llençaven papers a la professora i, seguidament, com, entre crits i burles, triaven al terra diverses taules i cadires. El vídeo anava acompanyat del següent text:

> "*Te mando un video de un centro educativo para los emigrantes menores de edad que entran ilegalmente en España.*
> *Te ruego que lo difundas para que España se entere de una vez como nos agradecen que los acojamos*"

Figura A.1.1. Captura de pantalla del vídeo que acompanyava el text publicat.

Per aquest motiu, la Secció de Delictes d'Odi i Discriminació Fiscalia Provincial de Barcelona, en la figura del seu Fiscal Coordinador, l'Il·lustríssim Miguel Ángel Aguilar García, va iniciar actuacions i, tal i com consta a les diligències d'investigació número 568/2019, la Unitat d'Investigació de Radicalismes a la Xarxa de la Comissaria General d'Informació dels Mossos d'Esquadra va poder identificar la persona que havia publicat el vídeo, una dona amb domicili a la província de Barcelona, i, a més, va comprovar que els fets que es mostraven al vídeo no havien succeït a Espanya sinó a Brasil i, per tant, no tenien res a veure amb el col·lectiu de MENAS.

Per aquests motius, la Fiscalia Provincial de Barcelona va considerar que existien indicis suficients de que els objectes descrits podien ser constitutius de delicte, consistent en la lesió de la dignitat de les persones per motius de discriminació relatius a la nació o origen nacional de l'**article 510.2.a) CP**, concorrent el subtipus agreujat de l'**article 510.3 CP** per la seva difusió a través d'Internet, concretament la xarxa social *Twitter,* i va presentar una querella criminal contra la dona que havia realitzat la piulada.

El dia 11 d'abril de 2023, la secció vint-i-unena de l'Audiència Provincial de Barcelona va dictar sentència de conformitat amb número de rotlle 79/2022 on es condemnava la dona a un any de presó i 900 euros de multa. Tanmateix, no s'ha pogut accedir al contingut íntegre de la sentència ja que, a causa de la proximitat temporal en la que va ser dictada, en la data en la que s'ha redactat el present annex aquesta encara no ha estat publicada a la web del Centre de Documentació Judicial del Consell General del Poder Judicial (CENDOJ)[82].

[82] https://www.poderjudicial.es/search/indexAN.jsp

ANNEX A.2. Primera condemna a Espanya per delicte d'incitació a l'odi per difondre una *fake news* a través de les xarxes socials

En aquesta ocasió ens trobem davant la primera condemna a Espanya per difondre una *fake news* emprant les xarxes socials, concretament la xarxa social *Twitter*, constitutiva d'un delicte d'incitació a l'odi de l'article **510.2.a) CP**.

Els fets queden recollits a la sentència de la secció sisena de l'Audiència Provincial de Barcelona, de 8 de novembre, amb número de registre SAP B 10887/2022 – ECLI:ES:APB:2022:10887[83], on es constata que la persona acusada, un home resident a la província de Barcelona, mogut per la seva animadversió i rebuig als immigrants estrangers d'origen marroquí i, més concretament, als menors que formen part del col·lectiu de MENAS, el dia 1 de juliol de l'any 2019, va publicar a la xarxa social *Twitter* un tuit on es podia veure un vídeo en el qual un home colpejava reiteradament i de forma brutal a una dona a la que posteriorment també semblava agredir sexualment. L'esmentat vídeo anava acompanyat del següent text:

> "*Aquí tenéis el video del MENA marroquí de Canet de Mar, a esos que le vamos a dar la paguita hasta los 23 años, los niños de Pedrito Piscinas. Por cierto, luego para más INRI la viola, estos energúmenos y estas manadas de marroquís no saldrán en los medios.*"

Figura A.2.1. Captura de pantalla del vídeo que acompanyava el text publicat.

[83] https://www.poderjudicial.es/search/AN/openDocument/41c846dea08d2cd1a0a8778d75e36f0d/20221121

La sentència constata que, amb aquesta publicació, l'home pretenia, amb un manifest menyspreu a la veritat, escampar de forma massiva i indiscriminada per la xarxa social *Twitter* una *fake news* en la que s'associava una agressió i presumpta violació, la qual posteriorment es va comprovar que realment havia succeït a Xina, amb el col·lectiu de menors d'edat estrangers no acompanyats. Tot això per desacreditar-los i difamar-los, atiant tot tipus de perjudicis i estereotips, amb la intenció de incitar o augmentar entre la població un sentiment de rebuig i hostilitat envers aquest col·lectiu. A més, la sentència també reflexa que el mateix home publicava de forma reiterada i indistinta als diversos perfils que tenia a les xarxes socials *Twitter, Facebook* i *Instagram,* missatges de caire supremacista, racista i xenòfob adreçats, en la seva majoria contra el mateix col·lectiu. Aquest fet va portar al Jutjat de Primera Instància i Instrucció 6 de Martorell, en virtut de les diligències prèvies 180/2020, de 4 d'abril, a acordar el tancament cautelar del seu perfil en la xarxa social *Twitter*.

En aquest sentit, la sentència considera provat que els fets detallats són constitutius d'un delicte de provació a l'odi i la discriminació de l'article **510.2.a) CP** en el que concorre el subtipus agreujat de l'article **510.3 CP**, per haver emprat un mitjà de comunicació social, en aquest cas la xarxa social *Twitter*, de tal manera que el contingut publicat va ser accessible per un elevat nombre de persones. També s'aplica l'article **510.5 CP**, amb penes d'inhabilitació especial per la professió o ofici educatius, i l'article **510.6 CP**, pel qual es decomissen i clausuren els perfils emprats per difondre missatges d'odi i discriminació.

Per tot l'exposat, la sentència, que, com en el cas anterior, és de conformitat, cosa que comporta que l'acusat reconeixia els fets i acceptava la pena imposada, condemna l'home a 15 mesos de presó, 9 mesos de multa amb una quota diària de 6 euros, 5 anys d'inhabilitació especial per la professió o ofici educatius, així com la clausura i decomís dels perfils a les xarxes socials assenyalats en els fets provats. Tanmateix, la mateixa sentència suspèn durant 2 anys la pena de presó si l'home no delinqueix en el transcurs d'aquest període, es compromet a mantenir clausurats els seus perfils i no n'obre de nous amb continguts discriminatoris i, finalment, participa en un programa d'igualtat de tracte i no discriminació establert pel Departament de Justícia.

ANNEX B. ÚS PRÀCTIC DEL PLUGIN INVID-WEVERIFY

Com ja s'ha explicat anteriorment, el plugin *InVID-WeVerify* és una eina **OSINT** específicament dissenyada per ajudar als verificadors de fets a combatre les *fake news* que es propaguen pel ciberespai. Aquest plugin disposa de quatre mòduls que ofereixen diverses eines per realitzar aquesta tasca. En aquest annex ens centrarem en dos d'aquests mòduls: el que ofereix eines per realitzar anàlisi d'imatges i el dedicat a analitzar vídeos. Per això, es mostraran dos exemples molt senzills però que alhora resulten molt clarificadors de com es realitza un procés de verificació d'imatges i vídeos.

ANNEX B.1. Anàlisi d'imatges

El mòdul d'anàlisi d'imatges del plugin *InVID-WeVerify* proporciona diverses per analitzar una imatge, ja sigui una de la que es disposa l'arxiu o una que es trobi en una xarxa social. Per aquest cas pràctic, emprarem una imatge que acompanya un tuit publicat a la xarxa social *Twitter* el dia 23 de maig de 2023. El tuit tracta sobra la invasió de camp per part d'aficionats del RCD Espanyol de Barcelona després del partit que va disputar aquest al seu camp contra el Fútbol Club Barcelona i on aquest darrer es va proclamar campió de Lliga, i afirma que en aquesta invasió diversos aficionats portaven bats de beisbol per agredir jugadors del Fútbol Club Barcelona. La publicació anava acompanyada d'una imatge on es podia veure com, suposadament, un home portava un bat de beisbol però, a més, una noia portava el que sembla ser una arma de foc, concretament un subfusell de guerra.

Figura B.1.1. Captura de pantalla del tuit publicat i la imatge que l'acompanya.

Per poder verificar si el que es mostra a la imatge es autèntica o ha estat manipulada es pot emprar, per començar, l'eina *Forensic*, que permet detectar falsificacions a les imatges mostrant una escala de colors concreta en el cas que detecti una manipulació. Com en el cas anterior, abans de res serà necessari carregar la imatge per poder-la analitzar.

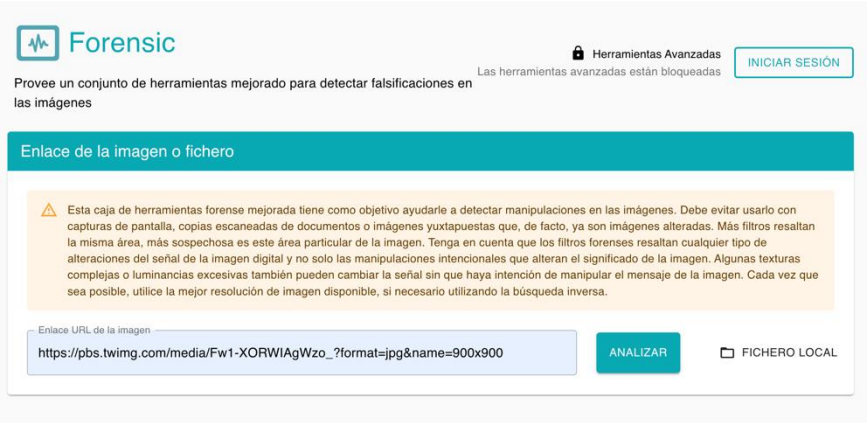


Figura B.1.2. Introducció de l'url de la imatge a analitzar.

En aquest cas, tal i com es veu a la següent imatge, l'escala de colors clars sobre les zones on es troben el bat de beisbol i l'arma és indicatiu que ambdós són elements que, amb molta probabilitat, poden haver estat afegits a la imatge.

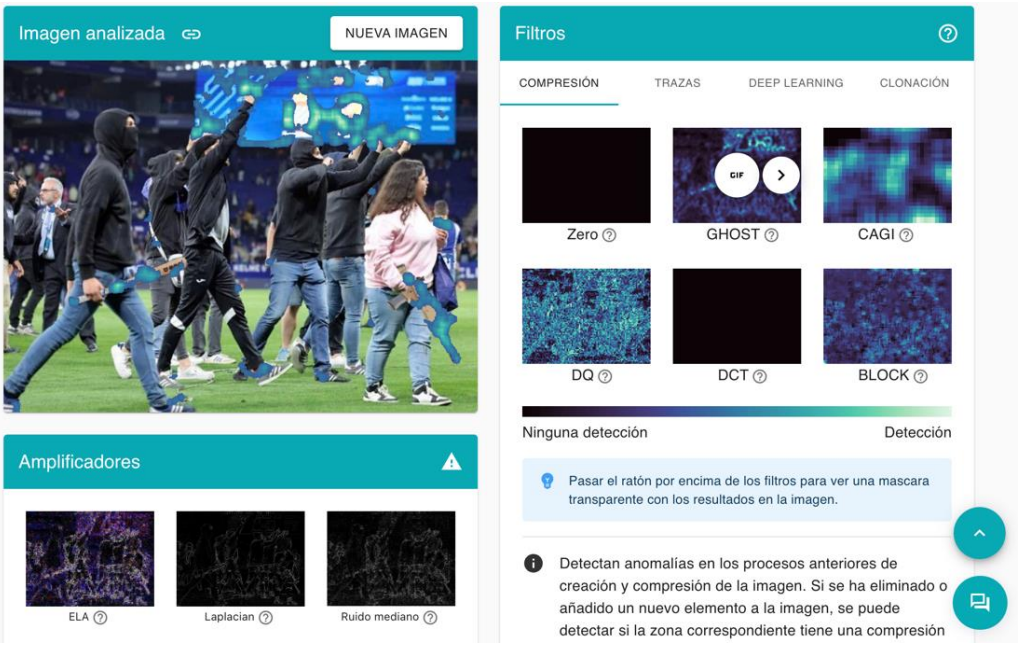


Figura B.1.3. Anàlisi de la imatge amb l'eina Forensic.

Una altra eina que ofereix el plugin és *Lupa*, que permet examinar la imatge amb deteniment per detectar diferències de textures i píxels que indiquin que es poden haver afegit o modificat parts de la imatge. Per això, primerament, caldrà facilitar la url (adreça web) on es troba penjada la imatge, en aquest cas, un perfil de la xarxa social *Twitter*.

Figura B.1.4. Introducció de l'url de la imatge a analitzar.

Un cop s'envia la imatge, ja es pot visualitzar i analitzar. En aquest cas interessa analitzar les parts on es veuen el bat de beisbol que suposadament porta el noi encaputxat i l'arma que, en teoria, sosté la noia del jersei rosa.

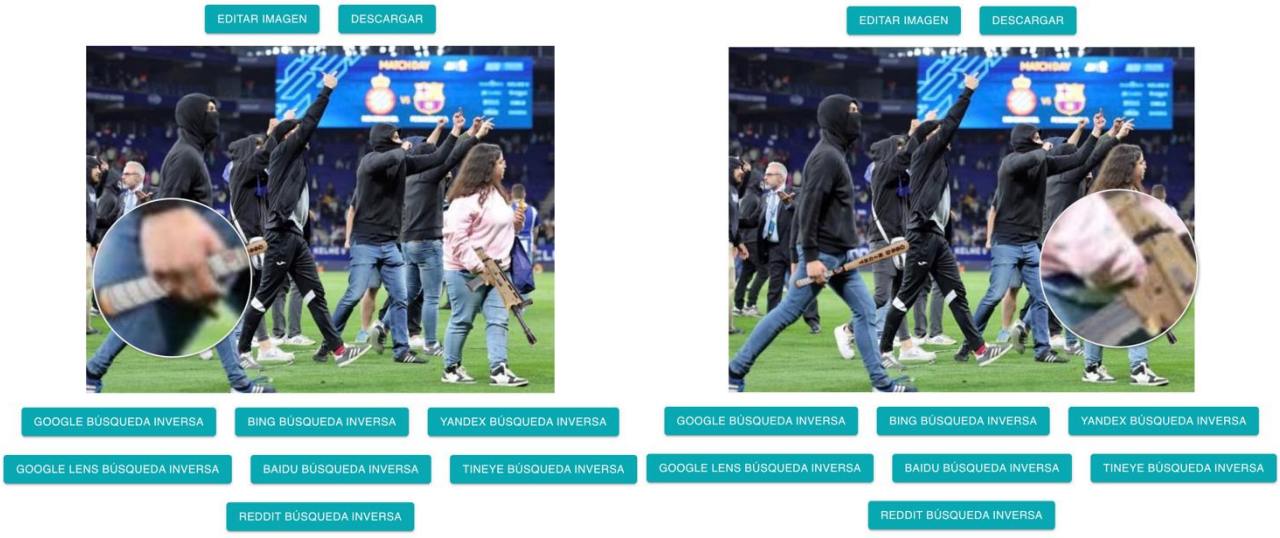


Figura B.1.5. Anàlisi de la imatge amb l'eina Lupa.

Com es pot veure en les imatges mostrades en la figura, a simple vista es pot apreciar com l'empunyadura del bat de beisbol sembla extreta d'un dibuix i la posició de la mà amb la que suposadament el subjecten no resulta natural. D'altra banda, posar la lupa per la zona on suposadament la jove porta l'arma, es pot veure clarament com no es possible que, tal i com està la mà de la noia, aquesta estigui realment agafant l'arma. Per tant, hi ha indicis que indiquen que la imatge ha estat editada per afegir el bat de beisbol i l'arma.

Finalment, una eina que pot resultar molt útil i eficaç a l'hora de detectar si una imatge és autèntica o ha estat manipulada a partir d'una imatge original és la recerca inversa d'imatges. En aquest cas, com es pot apreciar en la figura B.1.5, el plugin proporciona un total de set recursos que permeten realitzar aquesta recerca. Per aquest cas s'ha emprat *Google Lens*, a partir del qual s'han localitzat diverses pàgines web on surt la imatge original en la qual no surt ni el bat de beisbol ni l'arma.

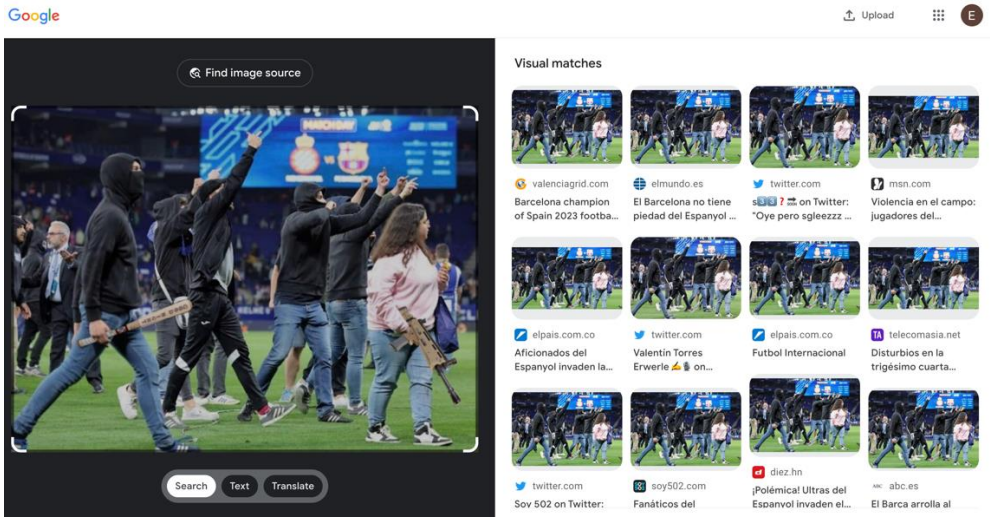


Figura B.1.6. Resultat de la recerca inversa d'imatges emprant *Google Lens*.

Per tant, mitjançant algunes de les eines que ens ofereix el plugin *InVID-WeVerify* s'ha pogut verificar de forma bastant senzilla que, efectivament, la imatge que acompanya el tuit analitzat ha estat manipulada per tal d'afegir-hi elements que no existien en la imatge original. És per això que es pot afirmar que el tuit és una *fake news*, ja que, si bé és cert que després del partit que entre el RCD Espanyol de Barcelona i Fútbol Club Barcelona hi va haver una invasió de camp per part d'alguns els espectadors, és totalment fals que alguns d'aquests saltessin amb bats de beisbol i armes per tal d'agredir als jugadors del Barcelona.

ANNEX B.2. Anàlisi de vídeos

Com en el cas anterior mòdul d'anàlisi de vídeos del plugin *InVID-WeVerify* proporciona diverses per analitzar un vídeo i així verificar si aquest és autèntic, ha patit algun tipus de manipulació o, simplement, s'ha emprat per donar suport a una notícia o informació que no té res a veure amb el que realment surt a les imatges enregistrades.

Per exemplificar aquest cas, i també per connectar-ho amb l'Annex A, s'ha utilitzat un vídeo en el que surt un home agredint de forma brutal a una dona i que va ser emprat per escampar una *fake news* en la que s'afirmava que l'home que sortia al vídeo era un MENA de Canet de Mar, tot i que posteriorment es va demostrar que, en realitat, les imatges que es surten al vídeo es corresponien amb uns fets ocorreguts a la Xina que res tenien a veure amb aquest col·lectiu.

Figura B.2.1. Tuit on es va penjar el vídeo atribuint falsament l'agressió a un MENA.

Com es dona el cas que, en aplicació de l'article **510.6 CP**, en la sentència judicial es va acordar la retirada del tuit així com el tancament del perfil des d'on es va publicar, no és possible accedir al vídeo original emprat. Tot i això, ha estat possible localitzar el vídeo en el canal de *Youtube* del diari "La Vanguardia", ja que el van penjar per explicar la notícia. És per això que, en aquest cas, només es farà servir una de les eines que ofereix el plugin per tal de demostrar que el vídeo, efectivament, reflexa uns fets ocorreguts a la Xina.

L'eina emprada és Fotogrames Clau, que, a partir de fotogrames del vídeo, permet realitzar una recerca inversa d'imatges, cosa que, tal i com es veurà, resultarà suficient per esbrinar-ne la seva procedència. Com en els casos anteriors, abans de res és necessari enviar la url del vídeo per tal que aquest pugui ser analitzat.

Figura B.2.2. Introducció de l'url del vídeo a analitzar.

Un cop enviat el vídeo, el plugin mostra una sèrie de fotogrames d'aquest, tal i com es veu en la següent imatge.

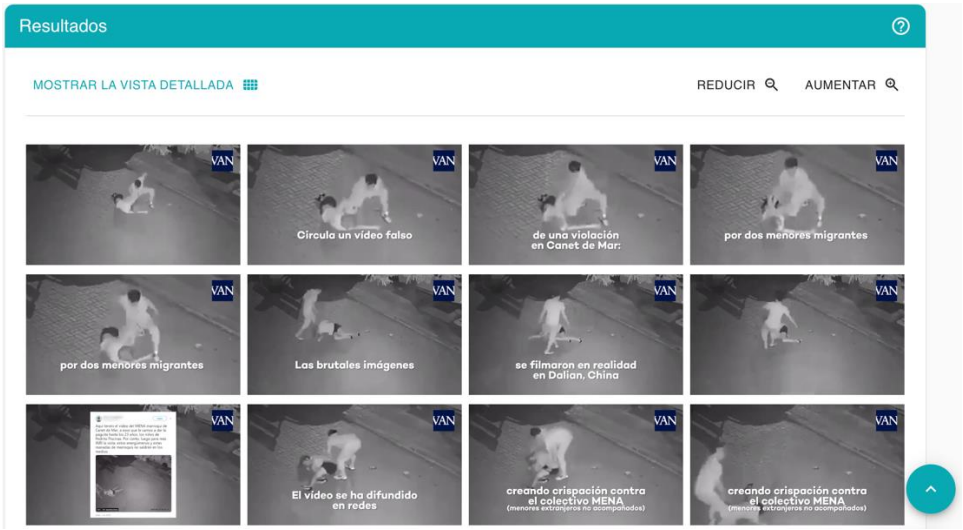


Figura B.2.3. Fotogrames obtinguts a partir del vídeo.

Seguidament, el plugin permet seleccionar una de les imatges i realitzar una recerca inversa emprant qualsevol de les opcions disponibles. En aquest cas s'ha emprat *Yandex*.

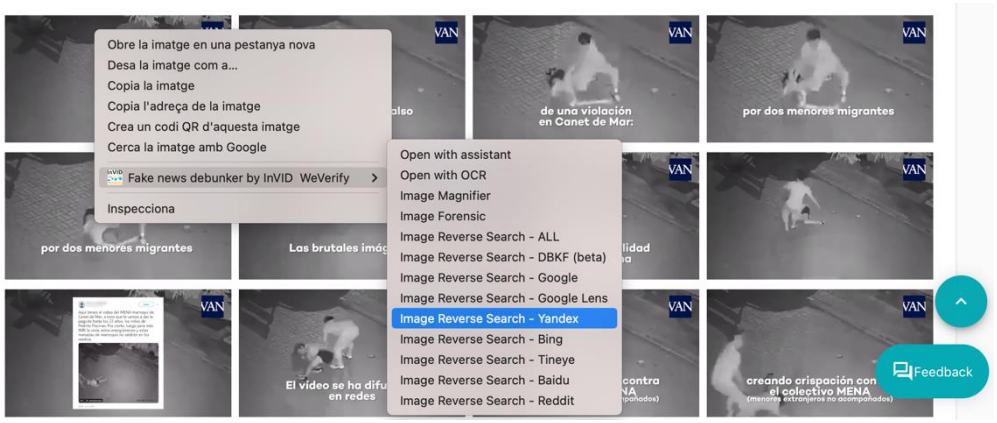


Figura B.2.4. Recerca inversa d'imatges emprant *Yandex*.

A partir de la recerca, s'obtenen una sèrie de resultats on es mostren imatges idèntiques o similars.

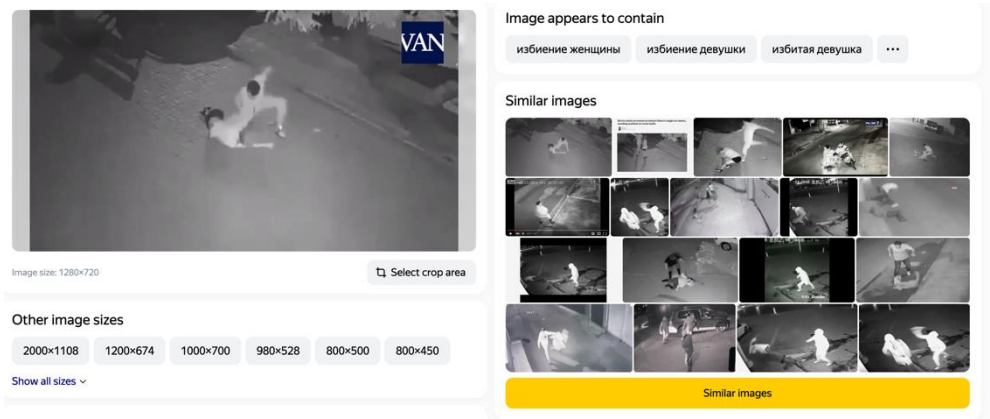


Figura B.2.5. Resultat de la recerca inversa d'imatges emprant *Yandex*.

A més, també es llisten aquelles pàgines web que contenen exactament la imatge emprada per realitzar la recerca. En aquest cas, entre el llistat de les pàgines web es pot observar que n'hi ha diverses de procedència xinesa. Accedint a una de les llistades, s'ha pogut arribar a una notícia d'un diari xinès on s'expliquen els fets que van propiciar l'agressió, ocorreguda a la ciutat de Dailan, i es mostra la verdadera procedència i identitat de l'agressor.

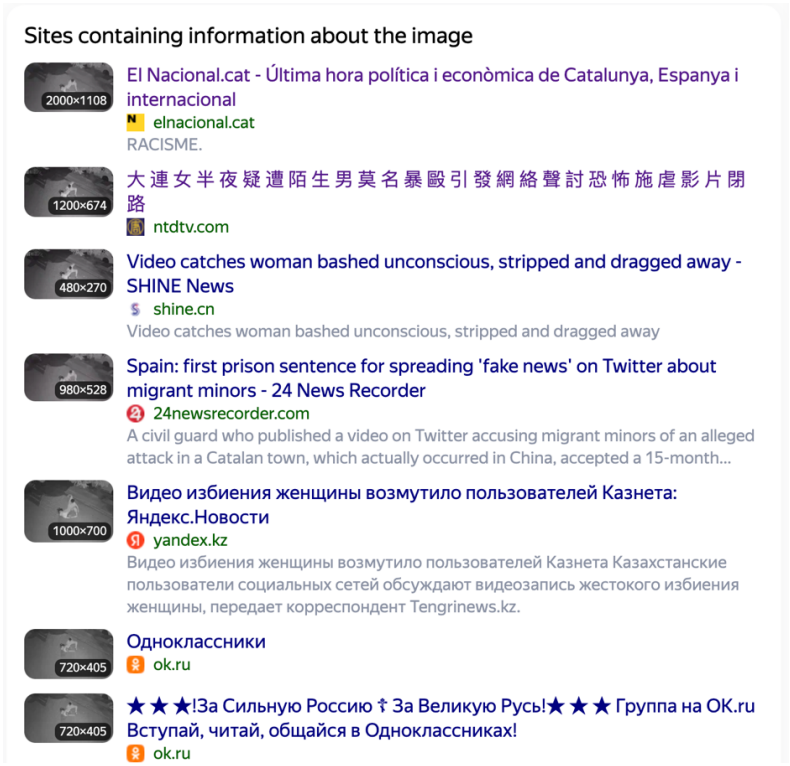
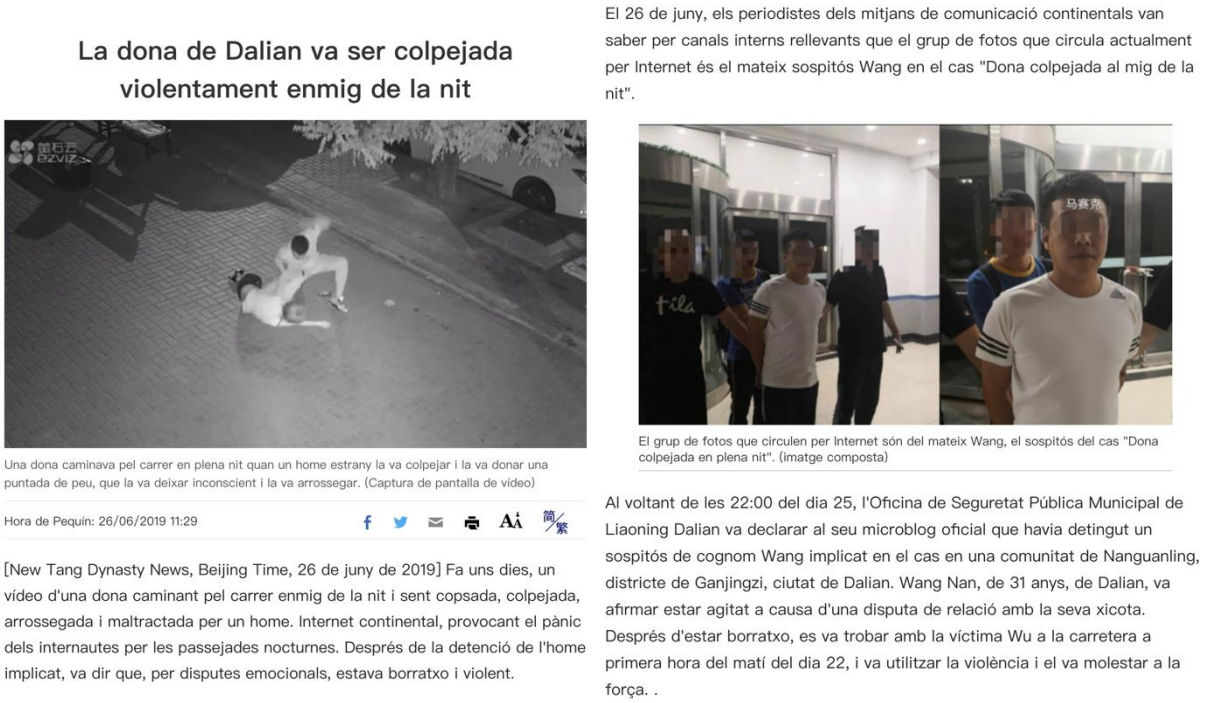


Figura B.2.6. Llistat de les webs que contenen imatges similars a la buscada.

Figura B.2.7. Pàgina web en xinès on s'explica la notícia de com van succeir els fets i com es va detenir l'agressor.



La dona de Dalian va ser colpejada violentament enmig de la nit

Una dona caminava pel carrer en plena nit quan un home estrany la va colpejar i la va donar una puntada de peu, que la va deixar inconscient i la va arrossegar. (Captura de pantalla de vídeo)

Hora de Pequín: 26/06/2019 11:29

[New Tang Dynasty News, Beijing Time, 26 de juny de 2019] Fa uns dies, un vídeo d'una dona caminant pel carrer enmig de la nit i sent copsada, colpejada, arrossegada i maltractada per un home. Internet continental, provocant el pànic dels internautes per les passejades nocturnes. Després de la detenció de l'home implicat, va dir que, per disputes emocionals, estava borratxo i violent.

El 26 de juny, els periodistes dels mitjans de comunicació continentals van saber per canals interns rellevants que el grup de fotos que circula actualment per Internet és el mateix sospitós Wang en el cas "Dona colpejada al mig de la nit".

El grup de fotos que circulen per Internet són del mateix Wang, el sospitós del cas "Dona colpejada en plena nit". (imatge composta)

Al voltant de les 22:00 del dia 25, l'Oficina de Seguretat Pública Municipal de Liaoning Dalian va declarar al seu microblog oficial que havia detingut un sospitós de cognom Wang implicat en el cas en una comunitat de Nanguanling, districte de Ganjingzi, ciutat de Dalian. Wang Nan, de 31 anys, de Dalian, va afirmar estar agitat a causa d'una disputa de relació amb la seva xicota. Després d'estar borratxo, es va trobar amb la víctima Wu a la carretera a primera hora del matí del dia 22, i va utilitzar la violència i el va molestar a la força. .

Figura B.2.8. Traducció al català de la pàgina anterior.

Per tant, com s'ha pot veure, emprant únicament una de les eines d'anàlisi de vídeos que proporciona el plugin *InVID-WeVerify* ha estat bastant ràpid i senzill trobar l'origen real del vídeo que va ser emprat per difondre la *fake news* sobre una suposada agressió per part d'un MENA a una dona a la localitat Canet de Mar. Recordar que la persona que va publicar aquesta fake news va acabar condemnada per un delicte de provació a l'odi i la discriminació de l'article **510.2.a) CP**.