EL PAPEL DE FACEBOOK, GOOGLE Y EL MALWARE EN LA DISTRIBUCIÓN DE FAKE NEWS: LOS ALGORITMOS Y LA ESCALADA TECNOLÓGICA COMO INFLUENCIA MEDIÁTICA

Daniel González Moreno

Tesis depositada en cumplimiento parcial de los requisitos para el grado de Doctor en

Investigación en Medios de Comunicación

Universidad Carlos III de Madrid

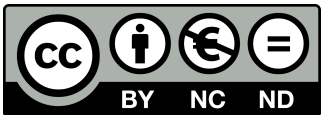
Director/a (es/as):

Carlos Elías Pérez

Tutor/a:

Carlos Elías Pérez

Mayo de 2023



AGRADECIMIENTOS

Esta tesis doctoral es un extenso trabajo que ha tomado años de esfuerzo y sacrificio que, de no ser por el apoyo de mis familiares, amigos cercanos y profesionales de la academia, no sería posible en absoluto. Por eso agradezco profundamente en primer lugar a mis padres por su confianza, amor y esfuerzo que han invertido durante toda su vida que me han permitido llegar hasta aquí. También a Virginia Moreno Echeverry, una de las grandes tutoras con las que tuve la fortuna de intercambiar ideas y recibir cientos de consejos en mis años de estudios de grado universitario. Y finalmente a Carlos Elías, mi director y principal apoyo en todo el proceso de estudio y desarrollo de esta tesis doctoral.

A todos ellos, muchas gracias por hacer esto posible.

ÍNDICE

INTRODUCCIÓN

Las *fake news,* y por consiguiente, la desinformación suponen en la actualidad uno de los retos más grandes a nivel periodístico, informativo e informático ya que pese a ser un fenómeno que no es nuevo han encontrado en los tiempos actuales una serie de herramientas tecnológicas que permite a sus autores escalar su alcance a niveles nunca antes vistos a través de las grandes compañías tecnológicas.

Los resultados y el éxito que alcanzaron las *fake news* en las elecciones presidenciales de los Estados Unidos en el año 2016 marcaron un antes y un después en la historia de la comunicación o, al menos, un punto histórico para las *fake news*, Alfonso *et al* (2019). Este episodio es uno de los más destacados en demostrar precisamente cómo las redes sociales suponen un recurso increíblemente útil para la desinformación.

La información falsa funciona y logra resultados políticos y económicos suficientes como para justificar su desarrollo y distribución a través de cualquier canal posible. Es aquí donde las personas e instituciones dedicadas a la distribución de información falsa lograron un hito histórico: distribuir contenidos a una escala sin precedentes a través de los canales de comunicación digital más populares: Meta -compañía que antes de 2021 era llamada Facebook[1]- y Google, el famoso duopolio de la publicidad digital, Cramer-Flood, E. (2021).

Esto es algo que es obvio en el marco de la teoría de la comunicación -el emisor y el receptor como partes necesarias en el proceso de comunicación- pero increíblemente novedoso en el ámbito del uso malintencionado de la comunicación en medios digitales.

Esto supone un hito precisamente porque Google y Meta cuentan con las audiencias más grandes a nivel global en distintos países e idiomas con las que logran entregar información personalizada gracias a su supremacía tecnológica, algoritmos de filtro y distribución de contenidos. No es para poco, con más de tres mil millones de usuarios en Facebook[2] -no confundir Facebook, la red social, con Meta, la compañía- y más de noventa mil millones de visitas al mes para Google[3] son los ecosistemas con mayor oportunidad de éxito para cualquier campaña de desinformación. Por primera vez en la historia la información falsa

[1] https://www.nytimes.com/2021/10/28/technology/facebook-meta-name-change.html
[2] https://about.fb.com/company-info/
[3] https://www.similarweb.com/website/google.com/#overview

puede llegar directamente a cada persona sin mayor intermediación y al mismo tiempo a gran escala.

Las grandes tecnológicas tienen una posición privilegiada en el mercado ya que ellas ofrecen herramientas verdaderamente útiles a la población en general, herramientas que son completamente gratuitas a cambio de recopilar información personal de cada uno de los usuarios. Esto les ha permitido cosechar con el paso del tiempo audiencias masivas que nunca habían sido concebidas en ningún medio de comunicación analógico.

No solo esto, sino que además de ser medios con un alcance mucho más amplio que el de cualquier otro, también ponen al alcance de cualquier persona una plataforma de comunicación con el que cualquiera puede lograr dicho cometido: ser escuchado por millones de personas.

Ahora, no se trata únicamente de una masa crítica de usuarios, sino también de la capacidad de analizar, almacenar y distribuir información a una escala nunca antes vista -y que crece constantemente-. Es decir, Google y Meta centralizan la distribución masiva de gran parte de la información que consumen las personas en la actualidad; no la producen, sino que la distribuyen.

Dicha posición de centralización hace que Google y Meta entiendan mejor que cualquier otra compañía qué tipo de contenidos son los que más disfruta o desea un usuario al usar sus servicios. Esto implica que quien conoce al algoritmo tiene la capacidad de llegar a una gran cantidad de personas. También implica que quien quiera tener visibilidad en cualquier servicio de estas tecnológicas tendrá que someterse precisamente a los "deseos" del algoritmo.

Millones de personas dan sus datos personales, fotos familiares e información de preferencias y gustos a diario a estas empresas, y ellas deciden qué contenido es el más apropiado para cada una.

Lo anterior genera una relación de dependencia cada vez más fuerte entre todos los actores involucrados en las dinámicas de comunicación modernas sin excluir por supuesto a las noticias. Los periódicos dependen de Google y de Facebook[4] y esto afecta la forma en la

[4] https://voxeurop.eu/es/los-medios-de-comunicacion-y-su-dependencia-de-google-y-facebook-un-problema-

que las noticias son redactadas y distribuidas. La realidad informativa de hoy es completamente distinta a la que se vivió en décadas pasadas no sólo por los cambios tecnológicos, sino también por los cambios en las dinámicas económicas que la tecnología trajo consigo.

En muchos casos se expandió y promovió la "democratización de la información" en donde ahora cada ciudadano tiene el poder de crear y distribuir información a su antojo, pero al mismo tiempo se creó una estructura monolítica y dependiente de "monopolios digitales" que son quienes definen qué se distribuye y de qué manera. Después de todo aunque todos tengan la capacidad de crear información, quien toma la decisión de cómo se va a ver, cuándo y de qué manera dicha información será Google, Meta, ByteDance u otra gran corporación.

Comprender esta realidad exige investigaciones desde distintos puntos de vista en donde varias disciplinas deben involucrarse para poder encontrar soluciones teóricas y prácticas al problema siempre desde la presunción de que nunca existirá una solución única e infalible, por el contrario, todo esto se trata de una escalada constante de mejoras, tecnologías y metodologías de verificación, detección y filtros para evitar la distribución de contenido desinformativo.

La investigación académica no puede separarse de la realidad económica, tecnológica y social que generan estos cambios y es menester de cualquier interesado realizar un análisis con estos elementos presentes en su perspectiva. Esto hace que sea necesario en el día a día de profesionales e investigadores de la comunicación deba existir un interés por comprender cómo las herramientas tecnológicas tienen una incidencia sobre la comunicación de las personas.

En otras palabras: está en manos de los investigadores, analistas e ingenieros lograr desarrollar soluciones cada vez más sofisticadas para garantizar el derecho a la información de los ciudadanos -considerando que la desinformación es una vulneración a ese derecho-.

Dicho proceso de comprensión exige en muchos casos una enorme capacidad de previsión para comprender cómo las grandes corrientes comunicacionales, tecnológicas y comerciales

---

para-la-libertad-del-periodismo/

convergen -o podrían converger- en el futuro. El mundo cambia demasiado rápido como para abordar únicamente el pasado, especialmente cuando los acelerados cambios experimentados en los últimos años han sido clave para la distribución de las *fake news*. En otras palabras: son precisamente los avances tecnológicos los que más han favorecido un entorno fértil para la desinformación -lo cual no quiere decir que este sea un fenómeno reciente-.

Desde distintos puntos de vista la distribución es considerada como uno de los principales factores de éxito de las *fake news* ya que sin importar cuán eficiente o engañoso pueda ser un mensaje, si este no es distribuído nunca será efectivo. A esto se suma la dificultad de profesionales e investigadores por comprender cómo funcionan los algoritmos que son precisamente responsables de la distribución de esta información ya que los gigantes tecnológicos -Google, Meta, ByteDance, etc.- son cajas de información opacas y no comparten mayor detalle sobre sus tecnologías.

Después de todo, los algoritmos y la forma en la que las grandes compañías tecnológicas desarrollan sus propias herramientas son los que crean retos y oportunidades en la distribución de información. Si no se llega a un punto de comprensión profunda de estas herramientas será imposible entender por completo cómo las *fake news* llegan a tener un lugar en las pantallas de los dispositivos de millones de personas.

Así mismo los autores de las *fake news* -sean empresas, organizaciones o personas- también se preocupan por entender cómo funcionan las herramientas tecnológicas responsables de distribuir toda la información que consumen los ciudadanos ya que esta es la mejor manera de alcanzar una audiencia considerable. “Si no entiendes a Google, a Facebook, a TikTok, a WhatsApp, etc. te será muy difícil alcanzar a millones de personas con tu mensaje”, esa es justamente la premisa con la que se vive actualmente.

Por lo anterior esta investigación tiene un enfoque técnico en donde se quiere explorar con precisión la distribución de las *fake news* específicamente en Google y Meta sin distinguir específicamente en un servicio en particular -Google cuenta con su servicio de buscador, la *Chrome Web Store*, servicios inactivos como *Google Now*, etc. y Meta incluye Facebook, WhatsApp, Instagram, *News Feed*, etc.-. Se busca entender a través de distintos capítulos cómo son distribuidas las *fake news*, qué cambios y evoluciones a nivel técnico existen en

los distintos servicios para detectar e impedir la distribución de contenido desinformativo, qué señales o información disponible actualmente podría estar siendo utilizada para calificar la veracidad de la información y finalmente qué métodos de distribución potencial pueden existir que aún no han sido considerados en el estado de la cuestión -o bien, tengan poca consideración o profundización-.

Esto además de ser un reto supone uno de los aspectos más interesantes de esta investigación ya que exige una dosis de creatividad e interpretación distinta a lo habitual ya que es necesario combinar información cuantitativa y cualitativa. Por ejemplo, un análisis de hechos verificables -el cambio del comportamiento de los periódicos en Facebook- contrastado con los comunicados de prensa de Meta.

En lo que respecta a las investigaciones referentes a las *fake news* en la actualidad, encontramos una cantidad considerable de éstas enfocadas en la detección y el análisis semántico, político y social sobre las *fake news*, es justamente este aspecto uno de los que llevó a la motivación investigativa para esta tesis, ya que pocos investigadores de las ciencias cuantitativas -ciencias de la computación- se enfocan en el análisis algorítmico de la distribución de las *fake news*; mientras que los investigadores más afines a las ciencias cualitativas -ciencias de la comunicación- no suelen incluir un análisis técnico o una reflexión tecnológica profunda en sus investigaciones que permita comprender la distribución de este tipo de contenidos.

No se descarta, por supuesto, la importancia del análisis semántico y de contenidos a la hora de estudiar y comprender el fenómeno de la desinformación e incluso hallar métodos eficientes para impedir su proliferación. La intención es simplemente llevar esta investigación hacia un enfoque que hasta el momento no es frecuente en la academia.

En otras palabras, esta tesis doctoral trata el tema de las *fake news* bajo una perspectiva relativamente escasa en la academia que, a su vez, es uno de los aspectos más importantes: la distribución. La diferencia principal entre otros estudios recae en que el estudio de la distribución está bajo un enfoque tecnológico.

También es interesante recabar sobre la relación que existe entre las *fake news* y la tecnología, no desde la óptica pura de la ingeniería informática, sino desde el punto de vista

de las ciencias de la comunicación; el motivo de ello está en la permeación que la tecnología consigue día a día en cada uno de los aspectos cotidianos de la humanidad, cosa que incluye sin lugar a dudas a los medios de comunicación. Es decir: las ciencias de la comunicación no pueden pretender conseguir un entendimiento pleno de las *fake news* sin una reflexión apegada a los avances tecnológicos y su impacto sobre la generación, el procesamiento, la distribución y el análisis del contenido.

Después de todo es la digitalización misma de la información la que ha conseguido transformar el texto, la imagen y el sonido en información matemática que puede ser interpretada por algoritmos.

La digitalización misma es parte integral de todos los nuevos procesos de comunicación y ello hace que su integración -aún desde una perspectiva no explícitamente técnica- sea necesaria en esta tesis además de futuras investigaciones.

Adicionalmente hay grandes tendencias tecnológicas con un ritmo de desarrollo vertiginoso en donde llama la atención de forma especial la Inteligencia Artificial, la cual facilita la generación de contenido en múltiples formatos. Esto seguramente traerá nuevos retos que sin lugar a dudas estarán involucrados con la desinformación y las *fake news.*

Y, en efecto, cada nueva tendencia tecnológica trae nuevos usos, escenarios y casuísticas que tienen que ser abordadas en el campo investigativo. Es evidente que la comunicación tiene una relación inseparable con la tecnología que, con sus ventajas y desventajas, es necesario comprender y observar ya que lo único que hará con el tiempo será crecer y complejizarse.

Por lo anterior uno de los objetivos de esta tesis es expandir la reflexión sobre la distribución de la información; Gran parte de las investigaciones se centran en el análisis cualitativo y cuantitativo de contenidos usando entre otras cosas técnicas avanzadas de *Machine Learning*, en la verificación de información y en la reputación de las fuentes.

Aunque estos enfoques investigativos tienen perfecto sentido y su valor científico es incuestionable es importante entender que comprender el cómo es distribuida la información es también un enfoque investigativo fundamental, después de todo han sido los nuevos canales de información digitales los que han dado tanto poder a las *fake news* en

nuestra historia reciente, además con el cambio constante de las herramientas de comunicación utilizadas hoy en día es evidente que surgirán métodos nuevos más eficientes para distribuir *fake news*. El conocimiento técnico de las plataformas y sus algoritmos son una arista más de todo el conocimiento necesario para combatir la desinformación.

Esto nos lleva a plantearnos algunas preguntas importantes:

- ¿Cómo logran distribuir las noticias falsas a través de Google y Meta?
- ¿Existen cambios importantes tras las elecciones presidenciales de 2016 en los Estados Unidos que afectasen la distribución de desinformación?
- ¿Es posible utilizar software malicioso para la distribución de *fake news*?

Responder a estas preguntas resulta ser un reto teórico y práctico en cuanto a que la investigación científica actual tiene una exploración relativamente poco profunda en lo que corresponde a la distribución de *fake news* específicamente en medios digitales y a su vez la mayor parte de fuentes fiables no son artículos académicos sino investigaciones periodísticas o notas de prensa.

Este es un escenario -que está detallado con el estado de la cuestión y el marco teórico- donde la primera necesidad es encontrar y organizar todo el conocimiento disponible en un solo lugar, es decir, es necesario que existan futuras investigaciones en el análisis de los algoritmos y en la experimentación sobre los mismos, así como la recopilación teórica de sus funciones y características.

La investigación académica se encuentra relativamente atrasada en comparación a disciplinas prácticas como el marketing digital, la cual utiliza los algoritmos a su favor con relativa facilidad gracias a los profesionales que comprenden su funcionamiento. Mientras que ya existen profesionales especializados en el análisis de los algoritmos de Google y de contenidos para posicionar contenido en el buscador y de otros especializados en los algoritmos de distribución de contenidos de Meta en la literatura académica existe poca profundización sobre la relación entre algoritmos y distribución de noticias falsas.

Por tanto el objetivo general -aparte de buscar una respuesta a las preguntas planteadas- es dar un primer paso para futuras investigaciones cuyo enfoque técnico sea un complemento

perfecto para el análisis de la distribución de contenido malicioso como pueden ser las *fake news*.

Durante el desarrollo de esta tesis veremos una división de distintos temas a través de capítulos que permitirán evaluar la distribución de las *fake news* a través de los medios digitales con un especial énfasis en Google y Meta.

Además de la distribución en sí misma serán evaluados aspectos fundamentales en dicho proceso de distribución ya que los algoritmos responsables de entregar un contenido específico a un usuario concreto necesitan de información detallada y por tanto les obliga a procesar información de manera constante.

Se descubrirá en cada uno de los capítulos de esta tesis cómo las *fake news* se enquistan socialmente ayudando a comprender cómo el rol protagónico de la tecnología afecta la forma en la que se piensa el contenido en la actualidad. Se vivirá una escalada tecnológica sin precedentes que exigirá una mayor transversalidad disciplinaria por parte de investigadores y periodistas.

Es entonces esta tesis el punto de partida ideal para los profesionales de las ciencias de la comunicación para comprender precisamente la relación entre tecnología, *fake news,* distribución y gigantes tecnológicos y sus efectos sociales. Se espera que este punto de partida inspire nuevas investigaciones que precisamente ayuden a perfeccionar y sobre todo a fomentar esta integración entre periodismo y tecnologías algorítmicas relacionadas con el contenido y la información.

Entre mejor conozcamos las herramientas responsables del funcionamiento de los servicios que las personas usan para informarse será mucho más fácil para comunicadores e investigadores el hallar nuevos temas de investigación, técnicas de detección y metodologías que detengan en mayor medida el éxito de las *fake news,* después de todo tras este fenómeno hay mucho más en juego que una red social o un partido político, son las mismas estructuras sociales y culturales, los cimientos de nuestra sociedad, los que se ven comprometidos a medida que la desinformación se hace un espacio mucho más difícil de erradicar en la sociedad. Esto en efecto involucra un nivel de responsabilidad social y de diversidad intelectual a la cual muchos no están acostumbrados.

Además de lo anterior se reconoce que los temas estudiados en esta tesis doctoral son objeto de una evolución notablemente rápida con una serie de tendencias aplicadas a la productividad que mejoran día tras día[5].

Esto implica que varias fuentes de información o referencias tecnológicas que puedan hacer referencia a la creación o detección de contenidos sean obsoletas con el paso de algunos meses o incluso semanas. Esto no significa que la información contenida en esta investigación no tenga valor, más bien indica que la perspectiva con la que debe valorarse la información contenida aquí es más la de una tendencia hacia futuro y no como una realidad palpable en el presente.

Para ejemplificar lo anterior podemos hablar de las tecnologías de generación de imágenes; durante 2022 se vio la apertura de DALL-E 2 al público en general[6] y Midjourney 4[7] solo semanas después del lanzamiento de Stable Diffusion[8]. Es decir, en menos de un año la sociedad ha experimentado el uso público de 3 grandes tecnologías generadoras de imágenes, algo difícil de imaginar en 2021.

También han surgido nuevas inteligencias artificiales con usos completamente innovadores como el de *Whisper*[9], una IA -Inteligencia Artificial- enfocada al reconocimiento de voz. También otras como *FakeCatcher*[10] la cual detecta *Deep Fakes*.

A lo anterior debemos sumar los grandes cambios que las grandes tecnológicas están sufriendo a partir de la crisis económica que se vive justo en el momento de escribir esta tesis doctoral.

Distintas empresas como Meta, Twitter y Amazon despiden miles de empleados[11] y buscan ser mucho más eficientes y competitivas. Esto sin duda tendrá influencia sobre su forma de trabajo y cambiará la forma en la que invierten en proyectos que aparentemente no generan

[5] https://dataconomy.com/2022/05/artificial-intelligence-trends-2022/
[6] https://shotkit.com/news/dall%C2%B7e-2-opens-to-the-public-you-too-can-generate-distorted-photos/
[7] https://www.midjourney.com/
[8] https://stability.ai/blog/stable-diffusion-announcement
[9] https://openai.com/blog/whisper/
[10] https://www.digitalinformationworld.com/2022/11/intel-introduced-fakecatcher-solution.html
[11] https://www.timesnownews.com/business-economy/after-meta-twitter-amazon-to-sack-10000-employees-wh a t-is-causing-the-firing-wave-in-tech-market-article-95537611

un ingreso adicional en sus estados contables -muy posiblemente los esfuerzos para combatir las *fake news* llegarán a mínimos durante 2023-.

Justo en una época de cambios tan profundos es difícil establecer qué va a ocurrir y por qué va a ocurrir, pero es perfectamente posible ver las tendencias generales que marcan el futuro de la comunicación en el ámbito digital, y es allí donde varios de los capítulos de esta tesis tienen lugar y valor.

De momento es posible concentrarse en 3 puntos importantes: 1. la tecnología mejorará y será más sofisticada; 2. será mucho más accesible al público; y 3. tendremos un aumento exponencial en cantidad de información. Sin más dilación, este viaje tecnológico y comprensivo da su inicio.

MARCO TEÓRICO Y ESTADO DE LA CUESTIÓN

Los académicos e investigadores no han logrado plantear una definición de "*fake news*" con la que exista un consenso general, existen multitud de definiciones que según cuál sea aceptada el tipo de contenidos que pueden ser catalogados como "*fake news*" varía drásticamente. Molina *et al* (2019) resumen las discrepancias académicas sobre la definición de *fake news* en tres puntos principales: la primera es si el contenido de sátira debería ser incluída en esta categorización, la segunda es si todas las "*fake news*" tienen que ser intencionales para ser categorizadas como tal y la tercera es si las "*fake news*" son una categorización binaria -es o no es- o si es una variable contínua -contenido que podría ser, por ejemplo, "definitivamente falso", "parcialmente falso", "completamente verídico", etc. Es decir, no tiene blancos y negros para clasificar las *fake news*, en su lugar es una gran área gris-.

Tener en cuenta estas discrepancias es importante ya que la posición tomada para definir lo que es y no es una *fake news* tiene el potencial de cambiar las conclusiones de esta tesis. Esto también nos obliga a hablar de otros conceptos cercanos a las *fake news* pero que no son *fake news* como tal: "*misinformation*" y "*disinformation*", dos palabras que pueden traducir al español como "desinformación" pero que en la literatura científica anglosajona contienen significados diferentes. Southwell *et al* (2017) proponen una diferenciación bastante concreta, mientras que ambos conceptos se refieren específicamente a información falsa, la diferencia está en la intención, "*misinformation*" sería información falsa sin que exista una intención específica por parte del autor para ello sino más bien una discrepancia entre distintas personas, mientras que "*disinformation*" se refiere a información falsa que existe tras una intención explícita para ello. Considerando esto de ahora en adelante está traducido el término "*misinformation*" como "información imprecisa" y "*disinformation*" como "información falsa". Aunque esto no constituye una traducción exacta de cada término sigue siendo la mejor forma de reflejar las definiciones de cada concepto y nos otorga al mismo tiempo un punto de partida inequívoco para la interpretación de cada término.

La definición del concepto de "*fake news*" tiene además un uso variado variado en el discurso público ocasionando que su definición sea controversial y una tarea compleja,

Tandoc *et al* (2021). Dicho uso además juega un rol importante en la categorización y definición de una *Fake New* ya que de ello depende el formato del contenido (contenido satírico, teoría conspiranoica, información fuera de contexto, etc.) y de las intenciones propias del autor, Tandoc *et al* (2018). En otras palabras: el uso popular del término "*fake news*" y la diversidad intrínseca en el formato y la intención de los autores responsables de los contenidos "desinformativos" e "imprecisos" traen un abanico de interpretaciones muy extenso que impiden dar una respuesta precisa para definir el concepto "*fake news*".

Para abordar la definición de *fake news* se ha optado por utilizar los conceptos de Gelfert (2018) quien define a las *fake news* de una forma muy precisa:

> "Fake news, I argue, is best defined as the deliberate presentation of (typically) false or misleading claims as news, where the claims are misleading by design. The phrase "by design" is then explicated in terms of systemic features of the process of news production and dissemination." [la presentación deliberada de afirmaciones (típicamente) falsas o engañosas como noticias, cuando las afirmaciones son engañosas por diseño" (Gelfert (2018), p. 85-86).

Conceptualmente la importancia de esta definición está en su especificación de que las *fake news* son producto de un diseño, es decir, las *fake news* no son un producto del accidente o del error humano, son un acto intencional -la intención del autor tiene que ser la de desinformar aunque esto sea algo difícil de verificar-. Esto permite diferenciarlas de la información imprecisa como puede ser el reportaje con información falsa debido a un error en los procesos de verificación de información, incluso del contenido humorístico que intenta lucir como una noticia real -ya que el contenido humorístico no tiene como intención el engaño, sino la sátira-. Así mismo esta definición no excluye al contenido que es factualmente veraz, es decir, un contenido que sea verdad, pero su exposición al público no tiene otra finalidad distinta a la manipulación -lo que las aleja de su fin periodístico-. Se puede entender en las palabras de Gelfert que la definición de *fake news* no es más que una línea que ayuda a diferenciarlo de otro contenido engañoso o falso, lo que cuenta es la intención del autor.

En este sentido es importante considerar que en la literatura científica disponible habrá autores y estudios que incluyan determinados contenidos que no son "*fake news*" en el

sentido estricto bajo el cual se acepta el término en esta tesis. En todo caso esta diferencia semántica es inevitable al no existir un claro consenso sobre este término.

Así mismo hay autores como Baptista y Gradim que recalcan características de las *fake news* como un método eficaz para definirlas como puede ser la intención de imitar portales noticiosos altamente confiables y reconocidos:

> *"Fake news intends to imitate the presentation of news and reports in order to acquire credibility and legitimacy. However, it is important to note that fake news is not limited to obtaining the appearance of the news. Fake news seeks to imitate journalistic writing, the way a news article includes a photograph, and the way a credible news site presents itself. One of the most popular fake stories in the United States was published on a fake website (abcnews.com.co) that closely resembles the ABC News website(abcnews.go.com), not only in format but also in name, in the logo, and in the networkaddress (URL). The similarity of fake news with the journalistic news reports enhances the seriousness of this menace regarding the other elements of disinformation. The reach or popularity that fake news achieves is intrinsically related to the way it is presented,since, under the guise of news, it obtains a public perception that gives it credibility."[Las noticias falsas pretenden imitar la presentación de las noticias e informes para adquirir credibilidad y legitimidad. Sin embargo, es importante señalar que las noticias falsas no se limitan a obtener la apariencia de la noticia. Las noticias falsas buscan imitar la redacción periodística, la forma en que un artículo de noticias incluye una fotografía y la forma en que un sitio de noticias creíble se presenta. Una de las noticias falsas más populares en Estados Unidos se publicó en un sitio web falso (abcnews.com.co) que se parece mucho al sitio web de ABC News(abcnews.go.com), no sólo en el formato sino también en el nombre, en el logotipo y en la dirección de la red (URL). La similitud de las fake news con las noticias periodísticas aumenta la gravedad de esta amenaza respecto a los demás elementos de desinformación. El alcance o la popularidad que alcanzan las noticias falsas está intrínsecamente relacionado con la forma en que se presentan, ya que,*

> *bajo la apariencia de noticias, obtienen una percepción pública que les da credibilidad.] (Baptista & Gradim (2022). p. 638)*

Esta característica expone la importancia que tienen las características visuales de las *fake news* ya que estas tienen que lucir reales, creíbles o, al menos, hacer pensar a los lectores que son desarrolladas por un autor confiable, y no es para menos, el engaño en muchas ocasiones puede ocurrir no por el contenido en sí mismo de la noticia, sino por la suplantación de un autor o de una entidad creíble.

Baptista y Gradim también ofrecen una definición muy concreta de lo que son las *fake news*:

> *"We define fake news as "a type of online disinformation (1), with (2) misleading and/or false statements that may or may not be associated with real events, (3) intentionally created to mislead and/or manipulate a public (4) specific or imagined, (5) through the appearance of a news format with an opportunistic structure (title, image, content) to attract the reader's attention, in order to obtain more clicks and shares and, therefore, greater advertising revenue and/or ideological gain"." [Definimos las fake news como "un tipo de desinformación online (1), con (2) afirmaciones engañosas y/o falsas que pueden o no estar asociadas a hechos reales, (3) creadas intencionalmente para engañar y/o manipular a un público (4) concreto o imaginario, (5) mediante la aparición de un formato de noticia con una estructura oportunista (título, imagen, contenido) para atraer la atención del lector, con el fin de obtener más clics y shares y, por tanto, mayores ingresos publicitarios y/o ganancia ideológica] (Baptista & Gradim (2022). p. 640)*

Esta definición en sí misma es interesante porque incluye dentro de ella las intenciones -o el objetivo- que las *fake news* suelen tener, eso además de mencionar a quiénes se dirige este contenido y con qué tipo de información. No obstante, esta definición presenta dos problemas a nivel conceptual al contrastarlo con la información detectada en esta investigación.

En primer lugar, al definir a las *fake news* como un contenido desinformativo "online" se reduce considerablemente el alcance que tiene este tipo de contenido en otros canales, además de cierta forma negaría la existencia del contenido desinformativo en tiempos anteriores a la era digital que se vive hoy en día. El segundo problema está en limitar el objetivo de las *fake news* en un propósito de monetización publicitaria cuando es evidente -en hechos conocidos como las elecciones presidenciales del año 2016 en Estados Unidos- que las intenciones detrás de las *fake news* pueden incluir fines políticos. Esto no implica que esta definición no tenga valor, todo lo contrario, es una definición muy útil para entender a nivel general el comportamiento de las *fake news* en la actualidad y es precisamente la existencia de estos dos problemas conceptuales -presentes en otras definiciones- las que llevan a esta investigación a acoplarse a la definición de Gelfert.

Otro aspecto a considerar es el estado actual de las *fake news* en la sociedad ya que, aunque este fenómeno es antiguo a nivel histórico, su uso generalizado y estudio detallado ha conseguido su auge sólo en años recientes -aspecto evidenciable en la Figura 3-. Esto lleva a autores como Elías *et al* (2021) a afirmar que precisamente las *fake news* aún experimentan cambios y un desarrollo vertiginoso que añade una capa más de complejidad en su definición. Después de todo, no es posible definir con precisión algo que aún experimenta cambios en su aplicación y uso. Justamente en el libro de Elías *et al* titulado "Manual de periodismo y verificación de noticias en la era de las fake news" cada uno de los autores se refieren a las *fake news* de una forma distinta.

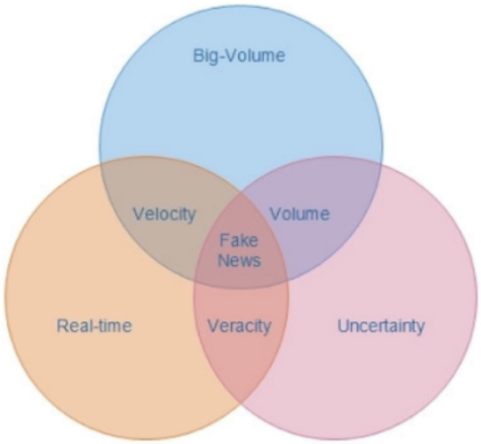
> *"¿Qué son entonces las fake news? Nuestros coautores se aproximan a su definición de modos muy distintos. Por ejemplo, una simple definición: información falsa presentada como verdadera (Fernández Roldán). O una tipología de 12 variantes de falsedades presuntamente informativas (García-Marín). O una clasificación según contexto, audiencia, narrativa y formato (Tuñón). No podemos dar una todavía una definición unificada, así que preferimos usar el anglicismo fake news mientras surge un consenso sobre en qué consiste el fenómeno." (p. 11)*

Coincidiendo con este punto de vista está claro que, incluso, definiendo una definición precisa en este marco teórico es imposible determinar futuros cambios que el entorno académico proponga en la definición de este anglicismo. Por ello lo apropiado es tomar la

definición seleccionada como un punto de referencia y no como una base universal que defina lo que es una *Fake New*.

Figura 1

Fake news en 3 vertices

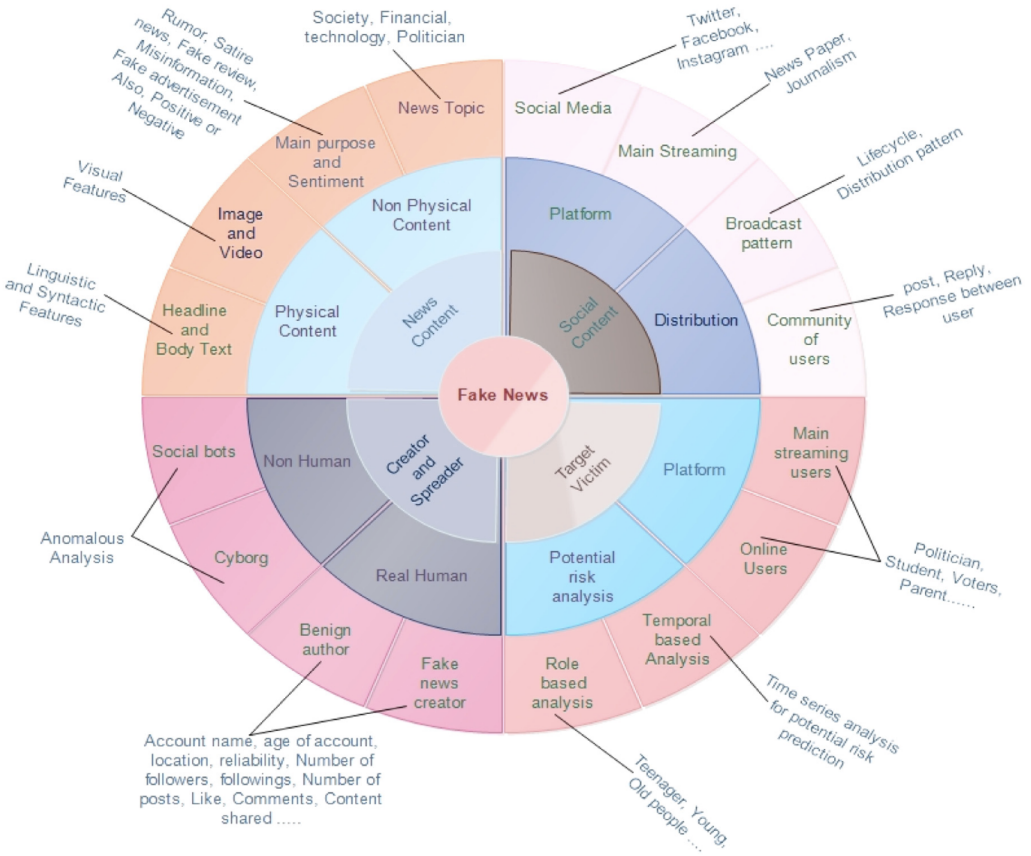


Fuente: Sahoo & Gupta (2021)

También debe ser considerado en la definición de las fake news características propias de su uso y distribución, por ejemplo, tal como expone la Figura 1, es necesario entender la presencia de las características propias del tiempo -tiempo real- y volumen como parte de las fake news. Es decir, una Fake New en sí misma debe competir con las noticias reales y verídicas y en tal caso también tienen que ser rápidas, abundantes y generar incertidumbre. Sin estas características las fake news, sencillamente, no tendrían lugar en la sociedad actual.

Figura 2

Características de las *fake news*



Fuente: Sahoo & Gupta (2021)

Sahoo & Gupta (2021) comparten un gráfico detallado sobre las características presentes en las *fake news* -Figura 2- en donde incluyen varias dimensiones que sintetizan adecuadamente la interpretación de lo que es una *fake news* ya que éstas dependen del contexto en el que están siendo desarrolladas y distribuidas, por esto vemos que los 4 ejes principales son precisamente el "contenido noticioso", el "contenido social", el "creador y el distribuidor" y finalmente "la víctima objetivo". Esto de alguna forma permite ver que

las *fake news* también cambian según su uso y contexto, después de todo no es lo mismo una *fake news* que tiene presuntamente objetivos comerciales -como puede ser una noticia falsa sobre el daño que hace el 5G sobre la salud humana- a otra cuyo objetivo es político -por ejemplo, decir que todos los inmigrantes son criminales-.

Varios autores mencionan la distribución del mensaje como una parte fundamental de la desinformación.

> *"Just as the invention of the printing press caused a surge in the proliferation of fake news, the advent of the Internet caused fake news to spread exponentially. Indeed, fake news reached its peak in 2016 with "Pizzagate." The general public was astonished to learn that a man shot open a locked door at a pizzeria in Washington, D.C. claiming to be investigating reports that Clinton aide John Podesta was heading up a child abuse ring in the parlor. The false political conspiracy theory claimed Hillary Clinton was coordinating a child trafficking ring at the pizzeria. At this point, society recognized the problem of fake news and began demanding a solution to prevent it." [Al igual que la invención de la imprenta provocó un aumento de la proliferación de noticias falsas, la llegada de Internet hizo que las noticias falsas se extendieran exponencialmente. De hecho, las noticias falsas alcanzaron su punto álgido en 2016 con el "Pizzagate". El público en general se asombró al enterarse de que un hombre abrió a tiros una puerta cerrada en una pizzería de Washington D.C. alegando que estaba investigando informes de que John Podesta, asesor de Clinton, dirigía una red de abuso de menores en el local. La falsa teoría de la conspiración política afirmaba que Hillary Clinton estaba coordinando una red de tráfico de niños en la pizzería. En ese momento, la sociedad reconoció el problema de las noticias falsas y empezó a exigir una solución para evitarlo.] (Watson (2018), parr. 17)*

Aunque probablemente la distribución del mensaje no sea una característica intrínseca de las *Fake New* es cierto que su existencia depende precisamente de esto -de la misma forma que el combustible no hace parte del coche, pero lo necesita para moverse-. Podría decirse en este punto que una *Fake New* efectiva, para ser considerada como tal, requiere de una exposición a un número considerable de personas.

Por qué existen las *fake news*

Las *fake news* son producto de un diseño ya que tras ellas existe un objetivo sea político, comercial o social, es decir, los autores -sean personas, empresas u organizaciones- tienen una finalidad que buscan cumplir a través de las *fake news* y precisamente para cumplir estos objetivos necesitan de una distribución masiva y eficiente de este contenido desinformativo. Después de todo un mensaje que no es visto por nadie no cumple con propósito alguno. Lo anterior hace que las *fake news* tengan un valor económico; autores como Kshetri & Voas (2017) proponen fórmulas para calcular el valor de las *fake news* las cuales incluyen variables como: beneficios monetarios, beneficios psicológicos, oportunidad de enganche con una audiencia, oportunidad de costo para la convicción de una audiencia, probabilidad de arresto y la probabilidad de convicción. Es decir, las *fake news* implican un grupo de riesgos y beneficios que se traducen en costos (producción y fiscalización) y beneficios (ingresos y cambio del pensamiento social).

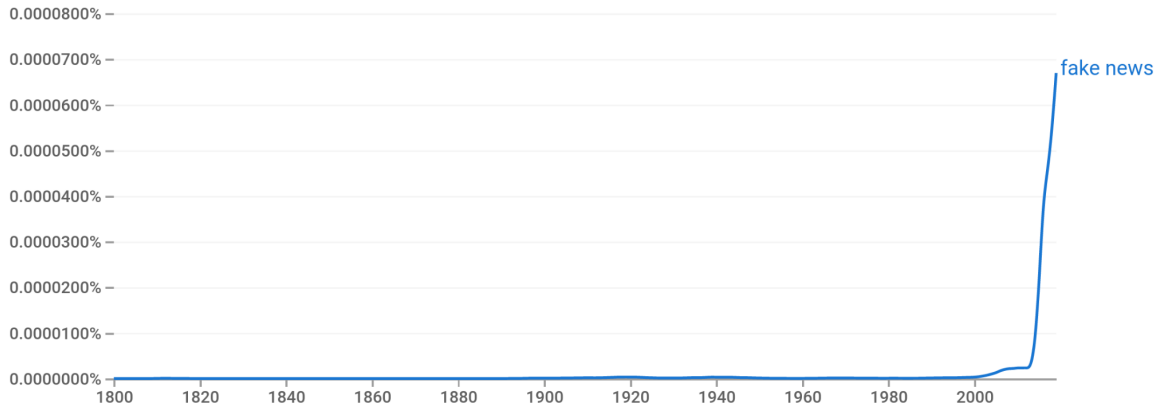
Nadie produciría una *fake news* si no tuviera algo que ganar y nadie invertiría tecnología, investigación y desarrollo en intentar combatirlas si éstas no pudieran transformar el actuar social, político y económico de una población determinada.

El rol de los autores de las *fake news* es hallar los métodos de publicación y distribución óptimos para aumentar ganancias mientras que el rol de los medios digitales -Google, Meta, TikTok, etc- es el de incrementar el coste de distribución de las *fake news* a tal punto que su distribución sea tan costosa que su incentivo sea mínimo. Esto es una escalada de armas tecnológicas con algoritmos de distribución de información en el medio.

Las noticias falsas o *fake news* son un tema de investigación llamativo en la comunidad científica contemporánea, en donde a pesar de tener un amplio campo de investigación por explorar ya hay un conocimiento de la importancia e impacto que suponen a nivel social y político, Landon-Murray *et al* (2019). Estos campos de investigación asociados a las *fake news* son diversos y los más populares se centran en las consecuencias de éstas a nivel psicológico, su impacto a nivel económico y político, su posible detección a partir de distintos métodos y técnicas informáticas, sus actores y medios de distribución principales, entre otros.

Figura 3

Ngram de la palabra "fake news" en Google Books



Fuente: Google Books Ngram Viewer[12].

Es interesante además ver que la presencia del término *fake news* en Google Books es un fenómeno relativamente reciente en donde a pesar de que su primer uso registrado en esta herramienta es en el año 1807 su uso es únicamente apreciable en la gráfica a partir del año 2001 con una clara tendencia al alza a partir del año 2013. Ciertamente, aunque las *fake news* son un fenómeno históricamente muy antiguo, su reflexión -al menos según Google Books- se ha hecho frecuente solo en la historia reciente.

Las motivaciones conocidas para la producción de una *fake news* son principalmente: beneficios económicos (monetización de sitios web y publicidad) y otros beneficios relacionados con el poder (política e ideologías), George *et al* (2021). Un excelente caso de estudio es el publicado por Massoglia (2020) quien detalla una compleja operación de inversión y dinero proveniente de la política norteamericana para financiar la creación y propagación de *fake news*, lo que demuestra el claro interés político detrás de toda la operación del contenido desinformativo. Pese a lo anterior aunque existe mucha información sobre la financiación de las *fake news* a través de la publicidad muy poco se sabe desde la academia sobre quiénes son los que financian de forma directa a los autores del contenido desinformativo. El anonimato que ofrece la red y las complejas redes empresariales son una clara barrera investigativa.

[12] https://books.google.com/ngrams/graph?content=fake+news&year_start=1800&year_end=2019&corpus=32&smoothing=50#

Un ejemplo perfecto es la investigación de Ahmed *et al* (2020) quienes investigaron el origen de la teoría conspiranoica que enlaza la tecnología 5G con el COVID-19 dando con la cuenta de Twitter *@5gcoronavirus19* como una de las más influyentes en el tema, no obstante no figura un origen específico en su estudio más allá de un clúster de cuentas, es decir: no se sabe quiénes estuvieron detrás de este suceso. Esto es particularmente llamativo al existir una coincidencia imposible de ignorar ya que el nacimiento de esta teoría encaja con las fricciones económicas y políticas por la implementación de la tecnología 5G de Huawei en todo el mundo[13]. Sin lugar a dudas el generar un malestar social y un rechazo directo a la tecnología 5G tuvo un efecto sobre la implementación de esta tecnología en algunos países como Suiza[14].

Las *fake news* asociadas al enlace entre el 5G y el COVID-19 son un ejemplo perfecto de dos aspectos: en primer lugar, los claros beneficios políticos y económicos que supone la distribución de información falsa en la población; y por otro lado lo difícil que es la detección de los autores y benefactores de toda la operación detrás de esta teoría conspiranoica. Es importante clarificar que, al no existir un conocimiento factual de los autores originales de esta teoría, la mera posibilidad de que alguna corporación u organización gubernamental estuviera detrás de este hecho es algo meramente hipotético; lo único que se resalta es que dicha teoría surge en un punto álgido de tensión política y económica con resultados que beneficiaron a una parte de dicho conflicto, un hecho imposible de ignorar.

Vale la pena mencionar que algunos autores recalcan que las *fake news* también pueden tener motivaciones altruistas o al menos no tener necesariamente una "intención maligna".

> *"Fake news was not solely written for financial benefits, however. Occasionally editors had altruistic motives. For example, in 1874, a New York Herald headline exclaimed, "Escaped Animals Roam Streets of Manhattan." The article described a startling mass escape of caged animals from the Central Park Zoo in alarming and graphic detail. It stated that as a result of the escape, twenty-seven people were dead and 200 individuals were savagely injured. The newspaper's editors had good*

[13] https://www.ft.com/content/90d53db6-ea7f-11e9-a240-3b065ef5fc55
[14] https://www.ft.com/content/848c5b44-4d7a-11ea-95a0-43d18ec715f5

> *intentions. The article was meant to draw attention to lax security measures at the Central Park Zoo. The final paragraph of the article spelled out in clear terms that no animals had actually escaped. Although the editors were resoundingly criticized by other newspapers for their scandalous tactics, the New York community did heed the warning and improved zoo conditions" [Sin embargo, las noticias falsas no se escribían únicamente para obtener beneficios económicos. En ocasiones, los editores tenían motivos altruistas. Por ejemplo, en 1874, un titular del New York Herald exclamaba: "Animales escapados vagan por las calles de Manhattan". El artículo describía una sorprendente fuga masiva de animales enjaulados del zoológico de Central Park con un detalle alarmante y gráfico. Decía que, como resultado de la fuga, veintisiete personas habían muerto y 200 habían resultado salvajemente heridas. Los editores del periódico tenían buenas intenciones. El artículo pretendía llamar la atención sobre las laxas medidas de seguridad del zoo de Central Park. El último párrafo del artículo explicaba claramente que ningún animal se había escapado. Aunque los editores fueron criticados rotundamente por otros periódicos por sus tácticas escandalosas, la comunidad neoyorquina hizo caso de la advertencia y mejoró las condiciones del zoológico] (Watson (2018), parr. 14)*

Aunque este tipo de motivaciones no sean las más frecuentes o incluso el ejemplo de Watson pueda no ser considerado como una *Fake New* al uso, es cierto que este tipo de casos ayudan a pensar en la desinformación con una óptima más extensa.

Siguiendo este hilo de pensamiento se concluye con la información disponible que producir *fake news* son desde la perspectiva del autor una clara inversión la cual tendrá su máximo beneficio solo a través de una gran distribución y por ende para entender a las *fake news* se necesita entender también su distribución, lo cual también implica un profundo entendimiento tecnológico al ser las tecnologías digitales los principales medios de distribución en la actualidad.

A pesar de que las *fake news* destacan por ser una herramienta de propaganda política también son un negocio lucrativo gracias a la publicidad digital; existen sitios dedicados a la distribución de *fake news* que generan ingresos a través de medios legítimos de publicidad digital como Google -es decir, empresas pagan a Google, y Google publica sus

anuncios en sitios que distribuyen *fake news*, generando ganancias- Papadogiannakis & Kourtellis (2022). Esto amplía la perspectiva con la que se analizan los incentivos del autor de las *fake news* ya que no se trata siempre de lograr un gran cambio social, en muchos casos los objetivos pueden ser tan simples como el de persuadir a la audiencia a través del Clickbait -textos llamativos que incentivan el clic en un enlace- para obtener tráfico y con ello ingresos publicitarios.

Además de las intenciones y objetivos del autor existen las intenciones propias del lector y de aquellos que distribuyen las noticias. Por ejemplo, para un militante -o seguidor- político no existe una gran diferencia entre compartir noticias reales o falsas mientras éstas sean útiles para denigrar al partido político opositor, Osmundsen *et al* (2021). Este es un matiz importante al momento de interpretar la propia distribución de las noticias falsas ya que las intenciones del receptor también juegan un papel fundamental en la eficacia necesaria para que la distribución de un contenido sea exitoso. No se puede entender una distribución exitosa de las *fake news* si detrás de ello no existe un objetivo tanto por parte del emisor como del receptor del mensaje.

Autores como Buchanan también hablan del rol que tienen los lectores en la distribución de las *fake news* al lograr perfilar a las personas responsables de distribuir o viralizar un contenido desinformativo:

> *"Typically, these will be people who think the material is likely to be true, or have beliefs consistent with it. They are likely to have previous familiarity with the materials. They are likely to be younger, male, and less educated. With respect to personality, it is possible that they will tend to be lower in Agreeableness and Conscientiousness, and higher in Extraversion and Neuroticism. With the exception of consistency and prior exposure, all of these effects are weak and may be inconsistent across different populations, platforms, and behaviours (deliberate v. innocuous sharing)." [Por lo general, estas serán personas que piensan que es probable que el material sea cierto o que tienen creencias consistentes con él. Es probable que tengan una familiaridad previa con los materiales. Es probable que sean más jóvenes, hombres y menos educados. Con respecto a la personalidad, es posible que tiendan a ser más bajos en Amabilidad y Conciencia, y más altos en*

> *Extraversión y Neuroticismo. Con la excepción de la consistencia y la exposición previa, todos estos efectos son débiles y pueden ser inconsistentes entre diferentes poblaciones, plataformas y comportamientos (intercambio deliberado versus inocuo).] (Buchanan (2020). p. 31)*

Ciertamente esto es un vector que transforma la manera en la que una *Fake New* es redactada y distribuida, ya que lo ideal para cualquier mensaje es llegar a las personas con las cuales tiene una mayor afinidad, y en el caso de las *fake news* serán aquellas con mayor posibilidad de "viralización".

Por lo anterior se concluye que la eficacia de las *fake news* no recae únicamente sobre el emisor del mensaje, las malas intenciones del receptor lo pueden convertir en un repetidor del mensaje para darle aún más poder de distribución y, además, una validación social al mensaje, haciéndolo aún más peligroso.

Talwar *et al* (2019) estudiaron precisamente el comportamiento del receptor al momento de compartir las *fake news* en las redes sociales y encontraron que el *FoMO -Fear of Missing Out-* y el *SMF -Social Media Fatigue-* están asociados positivamente con el compartir información falsa en redes sociales. Es decir, el estado social y psicológico de la sociedad también juega un papel fundamental en la distribución de las *fake news*. Sin un entorno saludable de interacción humana es más fácil que las emociones y la ausencia de verificación de información se normalicen en la población general. Es interesante que tanto el *FoMO* como el *SMF* son dos fenómenos asociados a los nuevos medios de comunicación -redes sociales- que existen gracias a los avances tecnológicos.

Alvi & Saraswat (2021) llegaron a conclusiones similares y además los resultados de su estudio sugieren que no existe una mala intención al momento de compartir *fake news* en redes sociales sino que las personas tienen una motivación al compartir información que consideran útil para los demás. Los usuarios sienten una satisfacción al momento de compartir noticias y este es el principal motivador para compartirlas, el problema está en que no existe un proceso de verificación de información durante esta dinámica digital.

Esto apunta a que las dinámicas sociales, la velocidad de consumo y la superficialidad con la que la sociedad consume un contenido son los principales factores que llevan a una

persona a compartir una noticia falsa. Dicha tendencia a compartir noticias falsas es mayor cuando una persona se encuentra afectada por estados como el *FoMO* y el *SMF*. Aquí el medio en sí mismo -las redes sociales- ofrecen, además de un poder de distribución gigantesco, una dinámica de consumo -compartir contenido e inmediatez de consumo- que facilita el éxito de las *fake news*. Está claro que cada vez hay una mayor saturación de información y cada vez hay una menor dedicación de tiempo a entender en profundidad lo la información consumida en línea.

Los sesgos cognitivos y la interpretación de la información

Como varios autores ya apuntan a lo largo de este marco teórico, las *fake news* también tienen una fuerte asociación con la psicología humana y en este sentido los sesgos cognitivos juegan un papel fundamental ya que son estos una característica humana que los autores de las *fake news* explotan con éxito. En este caso por sesgos se entiende como la poco idónea percepción y razonamiento que los humanos hacen sistemáticamente de la realidad, dichos sesgos generan una asimetría en la mente humana que le alejan del pensamiento crítico ya que la mente humana tiende a validar sus propias ideas y a ignorar aquellas pruebas que refuten o pongan a prueba sus propias conclusiones, Elías *et al* (2021).

Evidentemente en un mundo perfecto con una población sin sesgos y completamente educada las *fake news* tendrían el fracaso garantizado, pero precisamente el pensamiento irracional humano ocasionado por sus propios sesgos hace que el engaño y la desinformación tengan de forma inevitable un lugar en la comunicación.

Si bien pueden existir investigaciones y ejercicios cognitivos que muestran señales positivas en lo que a reducción del sesgo cognitivo se refiere, Massolo & Traversi (2021), a escala global este será un fenómeno del comportamiento humano que estará presente.

Ariely (2008) habla extensamente sobre los sesgos cognitivos y su rol en distintos comportamientos irracionales del ser humano, en donde podemos destacar varios ejemplos importantes:

1. Los humanos tienen una interpretación relativa de la realidad y por ello buscan puntos comparativos para tomar decisiones, en palabras de Ariely (2008) “la

mayoría de las personas no sabe lo que quiere a no ser que lo vea en un contexto." (p. 3).

2. Existe la llamada "coherencia arbitraria", el cual es un fenómeno que establece un punto de partida coherente en la mente de las personas. Ariely (2008) lo ejemplifica con un sistema de precios en donde expresa que "(...) una vez los precios se han establecido en nuestras mentes ello va a dar forma no solo a los precios actuales sino también a precios futuros." (p. 26). Esto implica que las ideas preconcebidas generan una disposición (positiva o negativa) frente a futuras decisiones. Las noticias falsas tienen el poder de moldear la llamada "coherencia arbitraria" de las personas.
3. Las expectativas afectan la interpretación y las experiencias que el cerebro procesa; si existe una idea previa de que algo será bueno, entonces esto probablemente lo será.

Tomando estas ideas concretas de Ariely es posible ver cómo los sesgos cognitivos suponen una oportunidad para las *fake news* ya que estos sesgos abren precisamente la posibilidad de que una información falsa pueda manipular el comportamiento humano. Más aún también suponen una línea de defensa importante para éstas ya que una vez una persona ha leído y está convencida de que una información es cierta -aunque sea información deliberadamente falsa- el solo hecho de alimentar su coherencia arbitraria hará que sea mucho más difícil convencerla de lo contrario. Un ejemplo perfecto serían los terraplanistas, estas son personas que en muchos casos son educadas y cuentan incluso con conocimientos en ingeniería y física, pero sin embargo llegan a estar convencidos de que la tierra es completamente plana. Este comportamiento no es más que uno de los cientos de ejemplos posibles que demuestran la existencia de los sesgos cognitivos.

Otros autores hablan del sesgo de confirmación en términos distintos pero que apuntan a la misma dirección:

> *"Confirmation bias, to which loss control professionals are vulnerable, encompasses seeking or interpreting evidence in ways that are partial to existing beliefs, expectations or a hypothesis already in mind. In musical terms, is it*

> *dissident or harmonious when you find that all injuries are the result of an unsafe act and an unsafe condition? In any injury situation, causal elements will be evaluated."[El sesgo de confirmación, al que son vulnerables los profesionales del control de pérdidas, consiste en buscar o interpretar las pruebas de forma parcial a las creencias y expectativas existentes o a una hipótesis que ya se tiene en mente. En términos musicales, ¿es disidente o armonioso cuando se considera que todas las lesiones son el resultado de un acto inseguro y de una condición insegura? En cualquier situación de lesión, se evaluarán los elementos causales.] Nickerson (1998)*

Nickerson menciona claramente cómo los sesgos de confirmación aplican en cierta medida a determinados profesionales; no obstante, este aspecto aplica también a la población en general, podría decirse que el sesgo de confirmación es un factor de la psicología humana que hace de las noticias falsas un elemento considerablemente llamativo en determinadas situaciones.

Por ejemplo, cuando una persona busca defender sus ideas podría estar sujeta a leer y estar de acuerdo con contenidos que se alineen precisamente con sus creencias ideológicas, aún cuando dicha información sea completamente falsa.

Es posible que, de hecho, sea este aspecto uno de los factores que más han permitido que las *fake news* estén presentes en temáticas donde la ideología juega un papel importante como lo es la política. Autores como Baptista *et al* (2021) concluyen que "The belief and dissemination of (fake) news are related to the political ideology of the participants, classified within the scope of the left–right political-ideological dimension" [La creencia y la difusión de las noticias (falsas) están relacionadas con la ideología política de los participantes, clasificada en el ámbito de la dimensión político-ideológica izquierda-derecha]. Aunque es cierto que Baptista *et al* mencionan en su investigación -centrada en Portugal- que son las personas con alineación política hacia la derecha los que tienen una mayor tendencia a creer y distribuir *fake news.*

> *"The fact that rightwing participants believe in pro-left fake news more than left-wing individuals contradicts confirmation bias and may suggest that the level of*

> *education and the age of individuals may interfere with the degree of acceptance of fake news. In fact, our results showed that the low level of education and the older age group had an influence on right-wing people in believing pro-left fake news." [El hecho de que los participantes de derechas crean en las noticias falsas pro-izquierda más que los individuos de izquierdas contradice el sesgo de confirmación y puede sugerir que el nivel de educación y la edad de los individuos pueden interferir en el grado de aceptación de las noticias falsas. De hecho, nuestros resultados mostraron que el bajo nivel educativo y la edad avanzada influyeron en las personas de derechas a la hora de creer en las noticias falsas favorables a la izquierda.] (Baptista et al, 2021, p. 12)*

Por supuesto es importante considerar justamente lo que mencionan los investigadores en sus conclusiones y es que es inusual -o al menos contradictorio con el sesgo de confirmación- que sean justamente las personas alineadas con la derecha política las que tiendan a creer con mayor facilidad en *fake news* que favorecen a la izquierda política. Según Baptista *et al* esto sugiere que es precisamente el nivel educativo uno de los factores que más influyen sobre la creencia en las *fake news*.

No obstante, y como punto a tener en cuenta es que creer en una *fake news* no implica o define la interpretación que una persona da sobre ésta; es decir: aunque una persona de ideología conservadora crea en una *fake news* de ideología progresista no significa que su interpretación de la noticia sea a favor del ala ideológica progresista. Con lo cual lo mencionado por Baptista *et al* no es necesariamente contradictorio con el sesgo cognitivo.

Aunque no sean contradictorias es posible encontrar otros autores que nuevamente asocian otros aspectos distintos al sesgo cognitivo como predominantes en la creencia y distribución de las *fake news.*

> *Across two studies with 3446 participants, we found consistent evidence that analytic thinking plays a role in how people judge the accuracy of fake news. Specifically, individuals who are more willing to think analytically when given a set of reasoning problems (i.e., two versions of the Cognitive Reflection Test) are less likely to erroneously think that fake news is accurate. Crucially, this was not driven by a general skepticism toward news media: More analytic individuals were, if*

anything, more likely to think that legitimate ("real") news was accurate. All of the real news stories that we used – unlike the fake ones – were factually accurate and came from mainstream sources. Thus, our evidence indicates that analytic thinking helps to accurately discern the truth in the context of news headlines. More analytic individuals were also better able to discern real from fake news regardless of their political ideology, and of whether the headline was Pro-Democrat, Pro-Republican, or politically neutral; and this relationship was robust to controlling for age, gender, and education. [En dos estudios con 3.446 participantes, encontramos pruebas consistentes de que el pensamiento analítico desempeña un papel en la forma en que las personas juzgan la exactitud de las noticias falsas. En concreto, los individuos que están más dispuestos a pensar de forma analítica cuando se les plantea un conjunto de problemas de razonamiento (es decir, dos versiones del Test de Reflexión Cognitiva) son menos propensos a pensar erróneamente que las noticias falsas son exactas. Lo más importante es que esto no se debe a un escepticismo general hacia los medios de comunicación: Los individuos más analíticos eran, en todo caso, más propensos a pensar que las noticias legítimas ("reales") eran precisas. Todas las noticias reales que utilizamos -a diferencia de las falsas- eran precisas y procedían de fuentes convencionales. Por tanto, nuestros datos indican que el pensamiento analítico ayuda a discernir con precisión la verdad en el contexto de los titulares de las noticias. Los individuos más analíticos también fueron más capaces de discernir las noticias reales de las falsas, independientemente de su ideología política y de si el titular era pro-demócrata, pro-republicano o políticamente neutral; y esta relación fue robusta al controlar la edad, el género y la educación.] (Pennycook & Rand, 2019, pp. 46-47)

Ciertamente el sesgo cognitivo no parece ser aquel factor que determina si una persona cree o no en una *fake news,* pero por lo menos sí un elemento que puede llevar a una persona a ser más abierta a consumir determinada información falsa. Pero, aunque el sesgo cognitivo no sea un factor tan notorio como el nivel educativo, sí es posible ver otras investigaciones en donde este aspecto psicológico tiene alguna implicación sobre lo que una persona puede creer.

Greene *et al* (2021) investigaron sobre los recuerdos de las personas y su asociación con las *fake news* y encontraron que la exposición a *fake news* puede hacer que las personas tengan recuerdos falsos especialmente si estos se asocian a las creencias ideológicas de cada uno. Esto ayuda a trazar la estrecha relación que el sesgo cognitivo tiene con las *fake news* ya que es un poco más extenso que el simple hecho de "creer en una noticia", también llega hacia aspectos más sensibles como la forma en la que la sociedad recuerda hechos concretos. Gracias al sesgo cognitivo las *fake news* pueden cambiar la forma en la que se ve y percibe la historia. En todo caso, estos investigadores apuntan a principios de comportamiento muy similares a los ya vistos.

> *"Participants with higher cognitive ability and analytical reasoning scores were less susceptible to false memories. Individuals with better knowledge about Brexit showed better discrimination between true and false stories, while self-reported engagement with the Brexit debate was associated with an increased tendency to "remember" any story, regardless of its truth. These results implicate a combination of social and individual factors in the development of false memories from fake news, and suggest that exposure to social identity threats may enhance the polarising effects of fake news." [Los participantes con mayor capacidad cognitiva y puntuaciones de razonamiento analítico eran menos susceptibles de tener falsos recuerdos. Los individuos con mejor conocimiento sobre el Brexit mostraron una mejor discriminación entre las historias verdaderas y falsas, mientras que el compromiso autodeclarado con el debate del Brexit se asoció con una mayor tendencia a "recordar" cualquier historia, independientemente de su veracidad. Estos resultados implican una combinación de factores sociales e individuales en el desarrollo de los falsos recuerdos de las noticias falsas, y sugieren que la exposición a las amenazas a la identidad social puede potenciar los efectos polarizadores de las noticias falsas.] (*Greene *et al, 2021*, *p. 1*)

Ciertamente los investigadores apuntan nuevamente a la habilidad cognitiva de los individuos como uno de los factores que moldean la tendencia a creer en noticias falsas -o en este caso a crear recuerdos falsos a partir de la desinformación-. El sesgo cognitivo no puede ser de ninguna manera el factor más influyente ya que la psiquis de una persona se

compone por multitud de factores culturales, sociales, económicos, académicos, etc. Es posible afirmar que el sesgo cognitivo es un hilo conductor y es el factor psicológico que permite alinear todas estas complejidades individuales hacia la posibilidad de creer en una información falsa.

Desarrollo de las *fake news*

Lo anterior hila perfectamente con la teoría moderna de la comunicación; la conceptualización del alcance de un mensaje no se limita únicamente al tamaño de su audiencia, sino también a su inmediatez. McLuhan (1996) ya hablaba sobre el impacto de las nuevas tecnologías en la comunicación -en su publicación original del año 1964, consultado en una edición del año 1996-:

> *Por supuesto, la automatización adopta el servomecanismo y el ordenador. Es decir, adopta la electricidad como almacén y acelerador de la información. Estas características de almacén, o «memoria», y de acelerador son esenciales en todo medio de comunicación. En el caso de la electricidad, lo que se almacena o se transporta no es una sustancia corpórea, sino percepción e información. En cuanto a la aceleración tecnológica, se está acercando ahora a la velocidad de la luz. Todos los medios no eléctricos no habían hecho sino apresurar un poco las cosas. La rueda, la carretera, el barco, el avión, e incluso el cohete espacial, carecen absolutamente de la cualidad de movimiento instantáneo.* McLuhan (Ed.). (1996) *(p. 356)*

Las ideas de McLuhan dejan claridad del factor acelerador que tiene la tecnología sobre la comunicación en sí misma. Tanto es así que es posible entender la historia de las *fake news* según distintos episodios históricos: la era pre-imprenta, la post-imprenta, la de medios masivos y por último la era de Internet; en donde se entiende que el comportamiento y la distribución de las *fake news* se adapta a los recursos tecnológicos de cada época, Burkhardt (2017). En otras palabras: la forma en la que los medios distribuyen el mensaje ha sido un factor decisivo sobre el rol y el impacto que las *fake news* han tenido en la sociedad. Es imposible entender a las *fake news* sin pasar por los métodos de distribución del mensaje.

De forma inequívoca la percepción misma de las *fake news* cambia gracias a la inmediatez que Google y Meta ofrecen en la actualidad. Prueba de esto está en la antigüedad misma de las *fake news*. Abad (2019) quien afirma que la primera *fake news* de la historia contemporánea data del año 1835 y que su impacto en la población proviene del mismo alcance y poder que los periódicos poseían en ese entonces; es decir, la distribución del mensaje sobre la audiencia fue clave en el mismo nacimiento de las *fake news* según Abad. Esto, obviamente, tiene otros matices históricos ya que la información falsa existe desde mucho antes; sin embargo, el punto de partida histórico propuesto por Abad es oportuno en el contexto de esta tesis al ser un estudio enfocado precisamente en los medios de comunicación modernos. Según esto es fácil asociar la masificación que Google y Meta (antes llamado Facebook) ofrecen en la actualidad a la mayor presencia de *fake news* en la actualidad; después de todo, si es más fácil distribuir un mensaje, será mucho más llamativo crearlo y distribuirlo en la red.

Lo anterior implica que el contexto actual ha sido propicio para que las *fake news* tengan un éxito contundente al momento de cumplir con sus objetivos. Dicho contexto incluye aspectos nunca antes vistos en la historia de la comunicación. Para empezar la humanidad nunca había contado con "medios de comunicación" tan gigantescos y poderosos. Por una parte, Meta, según su web en abril de 2019, tenía casi 2.400 millones de usuarios (en concreto, 2.380.000.000 usuarios) lo que le da una influencia líquida que trasciende fronteras políticas, ideológicas o físicas. Probablemente es el "*News Feed*" (el alimentador de contenidos de Facebook -un producto de Meta-) uno de los principales motivos que hacen de esta red social un fenómeno tan importante en la actualidad. Por otra Google es una de estas grandes tecnológicas cuyo buscador alcanza más de 83 mil millones de visitas al mes (SimilarWeb, 2020) siendo esta empresa la que controla qué aparece en su buscador y cómo aparece, por tanto es una de las compañías privadas que tiene el poder de moderar y limitar la distribución de las *fake news*.

Esto sin duda alguna moldea la interpretación del uso que dan estas compañías a sus algoritmos, que son los responsables de seleccionar y filtrar la información distribuida en sus servicios tecnológicos -motor de búsqueda y alimentador de contenidos-. Esto hace que su investigación y análisis sea tan llamativo. Ya hay autores como Giansiracusa (2021)

quien revela a través de una agrupación de fuentes cómo Google fue un factor determinante en los resultados electorales del 2016 en Estados Unidos, un aspecto en el que el Erich Schmidt -presidente ejecutivo de Alphabet- admitió como un hecho que requirió del trabajo y revisión de Google debido a su impacto en la política estadounidense[15]. Aparte de esto Giansiracusa también habla sobre cómo la sociedad moderna depende principalmente de las redes sociales para obtener información noticiosa en su día a día, y por tanto son los algoritmos de estas redes las responsables de informar o desinformar -según sea el caso- a las personas. Estos aspectos exponen cómo la tecnología es una parte fundamental de la desinformación.

El impacto que este duopolio de la publicidad tiene sobre la distribución de las *fake news* también ha sido mencionada por autores como Whittaker (2020) quien se refiere a Google y Facebook -ahora conocida como Meta- como el duopolio de la distribución al resaltar la facilidad con la que el contenido desinformativo es distribuido -y también dado de baja- por estas plataformas. De una forma u otra tienen un poder monopolístico.

El contexto contemporáneo combina distintos factores favorables para la proliferación y éxito de las *fake news*: una altísima escalabilidad en la distribución de información al contar con audiencias cuyo tamaño se mide en millones de personas, una inmediatez históricamente nunca vista y un control de distribución escaso al depender de pocos actores privados. Se Debe sumar también el rol de la "economía de la atención" como elemento transgresor en el comportamiento y funcionamiento de los medios tradicionales al ser el tamaño de las audiencias el elemento clave para delinear la viabilidad de un medio en el mercado como un espacio publicitario viable, Franck, G. (2019); Por tanto la atención es lo que determina el sostenimiento de los medios de comunicación, un modelo económico que lleva a los medios tradicionales a incluir *fake news* en su contenido ya que la creación de un contenido creíble y confiable requiere de un mayor esfuerzo en tiempo y dinero, Popiołek *et al* (2021). El raciocinio detrás de esto es bastante simple: el contenido más viral y llamativo es el que trae mayor rentabilidad y en contraparte el contenido más elaborado y riguroso el que menos. Los incentivos del mercado no facilitan la distribución del contenido

[15] https://www.fastcompany.com/40488115/alphabets-eric-schmidt-on-fake-news-russia-and-information-warfare

verificado por parte de los medios tradicionales, la gran mayoría de personas se deja persuadir por los titulares y dedican segundos a consumir un contenido llevándonos a un punto en el que la cantidad de tráfico que reciba un periódico es más importante que su rigurosidad al momento de garantizar su viabilidad económica.

Los cambios generacionales son también un aspecto a considerar en este panorama; estudios como el de Pérez-Escoda *et al* (2021) muestran que la generación Z en España, a pesar de tener un mayor consumo de contenidos en sitios web y redes sociales, son precisamente estos los medios en los que menos confían. De la misma forma los medios más confiables para ellos son la prensa y la radio, que a su vez son los que menos consumen. Una paradoja indiscutible.

Trninić *et al* (2022) hicieron un estudio sobre la percepción de las *fake news* en distintos grupos generacionales encontrando que los más jóvenes pese a ser conscientes de la existencia del contenido desinformativo su análisis del contenido es superficial y carente de criterio. En contraste los adultos de mediana edad presentaron en este estudio una posición mucho más crítica y con un análisis mucho más profundo sobre la información que consumen.

Lo anterior nos permite evidenciar algunos aspectos fundamentales sobre la realidad de las *fake news* en la actualidad: las generaciones más jóvenes consumen contenido más digitalizado, mucho más inmediato y diverso llevándolos a un gran volumen de información con un nivel de desarrollo y análisis muy superfluo, el cual a su vez ellos tienden a no profundizarlo a través de otros medios. Aunque los medios tradicionales aún cuentan con un nivel notorio de confianza es cierto que el consumo de medios alternativos en la población erosiona cada vez más estos niveles de confianza. Finalmente, los medios tradicionales compiten por las audiencias que los medios digitales -redes sociales- han ganado con el tiempo, lo que les ha obligado a invertir menos en calidad de la información -otro aspecto que lleva a erosionar su confianza-.

Fake news y tecnología

Para desarrollar la relación entre *fake news* y tecnología hay que considerar que ésta se da en distintos niveles ya que la tecnología facilita la distribución del mensaje, otorga espacios propicios para la existencia de cámaras de eco, tiene herramientas perfectas para la elaboración de contenido a escala y difícil de verificar para la población en general y crea espacios que proporcionan el anonimato. En contraparte permite estudiar mejor el contenido desinformativo, facilita la detección de las *fake news* y otorga información oportuna para comprender su impacto. Para entender mejor este principio de relación tan complejo hay que ver el contexto general que relaciona ambos conceptos el cual luce como una constante "escalada de armas".

Los lazos entre *fake news* y tecnología tienen todo tipo de efectos en la sociedad. Ralston *et al* (2018) concluyeron en su estudio que la propagación de las *fake news* impacta significativamente sobre la percepción de la integridad de los sistemas de comunicación -como pueden ser los periódicos- lo cual abre lugar al escepticismo, la divergencia y la intolerancia. Visto desde afuera estas conclusiones nos ayudan a detectar un círculo vicioso en la comunicación moderna, en donde las nuevas tecnologías propagan contenido desinformativo, el cual incentiva la desconfianza en los medios tradicionales, lo que a su vez sigue incrementando el uso de las redes sociales y los medios digitales para consumir *fake news* -que la población en general llega a considerar como fuentes fiables de información-. Este fenómeno es observable en el artículo de Silverman (2016) en donde las interacciones en redes sociales para las *fake news* superaron las interacciones de las noticias de fuentes noticiosas tradicionales. Es precisamente la distribución de las *fake news* en Google, Meta, Telegram, WhatsApp, entre otros, lo que sigue empujando el consumo en estos mismos canales. La capacidad tecnológica de los algoritmos, su inmediatez y conveniencia son la punta de lanza de las *fake news*. Al momento de escribir esta tesis doctoral también se experimentaron hechos empíricos que demuestran que tal impacto tiene la tecnología sobre la distribución de la información que Rusia ha decidido bloquear a Facebook en su territorio en plena invasión en Ucrania, Milmo (2022); obviamente Rusia no haría esto si Facebook no tuviese la capacidad de incidir en la opinión pública. Esto hace

lógico que las *fake news* tengan una distribución mucho más rápida y fácil a través de los medios sociales, Shu *et al* (2017).

Esto nuevamente está conectado con la economía de la atención, que es lo que precisamente ha permitido a las grandes plataformas tecnológicas llegar a este punto al utilizar algoritmos adaptativos que son refinados constantemente con el fin de generar una adicción en los usuarios, Bhargava & Velasquez (2021). Generar este efecto de adicción en los usuarios aumenta sus tiempos de permanencia y de consumo de contenidos, que es precisamente lo que explica las grandes audiencias que ciertas las redes sociales tienen hoy en día. Es decir, el incentivo de Meta -la compañía a cargo de Facebook- y Google no es el de fomentar un espacio de análisis e información rigurosa, su incentivo real es el de diseñar espacios digitales adictivos en donde sus usuarios disfruten del contenido disponible.

Otro efecto a considerar es la posición monopolística que estas plataformas han alcanzado. Por ejemplo, Google utiliza todo su poder de mercado -al ser el buscador más usado y una de las compañías con mayor presupuesto a nivel global- con el fin de eliminar cualquier competencia real o potencial, Smyrnaios (2019). Varias de estas prácticas ejercidas por Google son la inclusión forzada de su propio contenido en la parte superior de los resultados de búsqueda o eliminar determinados sitios web de sus resultados de búsqueda para reducir los ingresos económicos de sus competidores, Mays (2015). Lo evidente es que si una compañía permanece en una posición monopolística los usuarios no tendrán otra opción más que continuar usando sus servicios, alimentando aún más este fenómeno de "grandes audiencias" que ha sido mencionado.

Las grandes tecnológicas tienen una gran concentración de poder que está materializado a través de audiencias gigantescas, y eso tiene efectos en la población y en el mercado que está reflejado en los nuevos hábitos de consumo: más dispositivos digitales, menos medios analógicos -TV, Radio, periódicos, etc.-

Törnberg (2018) habla sobre la distribución de las *fake news* en las redes sociales y expone que los distintos factores del mundo real, la segregación de usuarios y las complejas interacciones en la red pueden jugar un papel fundamental en la difusión de información. Esto implica que la distribución de las *fake news* y su impacto no se limita únicamente por

lo que ocurre al interior de las redes sociales y que, de hecho, factores externos -como un dispositivo infectado por un software malicioso o incluso el voz a voz que las mismas personas practican entre sí- pueden ser un elemento a tener en cuenta al tener el potencial de segregar usuarios al exponerlos a contenido desinformativo. El mensaje no se detiene en un post.

Es entonces la tecnología un potenciador del alcance y la distribución del mensaje y por tanto es esta misma la principal herramienta con la capacidad de permitir a las *fake news* cumplir con este primer propósito: llegar a millones de personas en tiempos muy cortos. Esto no solo está reducido a lo que Google y Meta pueden hacer como plataformas, esto se extiende a cualquier plataforma digital, es decir, todo aquella tecnología que facilite la comunicación inmediata y la transferencia a escala de cualquier tipo de información ya está posicionada como un posible canal de distribución propicio para las *fake news*. Además de ello cada tecnología contiene sus propias características y dinámicas que a su vez afectan el formato mismo de las *fake news* -en Facebook un grupo de usuarios es perfecto para distribuir *fake news* a través de publicaciones, mientras que en TikTok difundir videos de corta duración fáciles de viralizar-. Las *fake news* usan las tecnologías para lograr impacto y distribución, mientras que la tecnología moldea la forma y el formato de las *fake news.* Un gran ejemplo de esta realidad está en España en donde WhatsApp se convirtió en uno de los medios alternativos de comunicación más importantes durante el confinamiento del año 2020 a raíz del COVID-19 a raíz de una creciente desconfianza en los medios de comunicación tradicionales, Elías & Catalan-Matamoros (2020). Esto obviamente refleja otro tipo de fenómenos como el constante crecimiento del consumo de medios digitales por encima de la televisión[16] pero sigue siendo innegable cómo la tecnología en sí misma transforma la forma de distribución de un mensaje, transformación que los autores de las *fake news* aprovechan frente a la más mínima oportunidad.

La tecnología tiene tal poder sobre el mensaje que también se ha convertido en una potente herramienta para la producción misma de noticias falsas; es decir, los últimos avances tecnológicos también son usados para crear mensajes mucho más realistas, convincentes y a

[16] https://www.lavanguardia.com/television/20220502/8236854/mes-abril-sido-menor-consumo-televisivo-historia-pmv.html

un ritmo nunca antes visto; la tecnología no está limitada a la mera distribución. Por ejemplo las *Deep Fakes* -un contenido de audio o video creado con algoritmos de deep-learning que logran un resultado realista de personas reales diciendo cosas que nunca dijeron, Chesney & Citron.(2018)- son una herramienta con el poder de persuadir de una forma efectiva a un grupo de votantes, Paterson & Hanley (2020). La tecnología convierte a las *fake news* en armas electorales, armas cuya creación está al alcance de cualquier persona con un ordenador.

La generación de contenido usando únicamente *Machine Learning* -u otras técnicas de inteligencia artificial- son un aspecto muy extenso y además preocupante. Giansiracusa (2021) explica a través de múltiples ejemplos que las computadoras ya tienen el poder de generar titulares, artículos periodísticos, perfiles falsos en redes sociales, entre otros; uno de ellos data la existencia de múltiples autores falsos en sitios de noticias falsas intentando llevar una narrativa política específica[17].

Grandes avances tecnológicos al alcance de toda la sociedad

El poder de estos algoritmos es fácil de materializar a través de múltiples sitios web, un gran exponente de ello es https://thispersondoesnotexist.com/, un sitio web con rostros humanos de personas que no existen ya que estos rostros fueron utilizados con inteligencia artificial -concretamente con una red generativa antagónica llamada StyleGAN[18]. Si los ciudadanos pueden acceder a estos recursos con tanta facilidad, de forma gratuita y además sin ningún conocimiento técnico avanzado es fácil entender que estas mismas herramientas pueden ser utilizadas para propósitos poco éticos.

El hecho de poder crear con gran facilidad distintas imágenes que parecen ser reales suponen una gran ventaja para los autores de las *fake news;* tal como explican Elías *et al* (2021) sobre el factor de credibilidad que tiene una pieza de video en Internet en el cual la percepción de la calidad es un factor que potencia la credibilidad de un mensaje. Si cada vez es mucho más fácil crear material de alta calidad que difícilmente es distinguible de un

[17] https://www.thedailybeast.com/right-wing-media-outlets-duped-by-a-middle-east-propaganda-campaign
[18] https://es.wikipedia.org/wiki/StyleGAN

material verídico es evidente que estas herramientas tienen un potencial vigente de engaño y manipulación mediática.

¿Es posible construir una fotografía falsa?

Tras la salida de Stable Diffusion en 2022 surgieron distintos debates con cuestiones legítimas sobre el uso de esta tecnología como podría ser su uso ético[19] o incluso legal[20]. Todas estas cuestiones además de ser perfectamente válidas necesitan de ejemplos reales aún si son solo teóricos. Para ello en esta investigación se propuso la generación de un contenido considerado como problemático.

Figura 4

[19] https://techcrunch.com/2022/08/24/deepfakes-for-all-uncensored-ai-art-model-prompts-ethics-questions/
[20] https://www.theverge.com/2023/1/16/23557098/generative-ai-art-copyright-legal-lawsuit-stable-diffusion-midjourney-deviantart

Fotografía de Hillary Clinton en la cárcel generada con Stable Diffusion -Inteligencia Artificial-

Fuente: Elaboración propia utilizando el prompt "photography of hillary clinton in the jail wearing jail clothing, sad, concerned, jail clothing, trial".

Un ejemplo perfecto del uso de esta tecnología es la posibilidad de crear una fotografía falsa con el objetivo de promover una falsa narrativa. La figura 4 es una imagen completamente falsa que simula una fotografía de Hillary Clinton en la cárcel. Por increíble que parezca esta imagen no es un fotomontaje, no es una alteración fotográfica, es simplemente una imagen generada por completo utilizando inteligencia artificial.

Estas herramientas son accesibles para cualquier persona, un ejemplo perfecto son las *Deep Fakes* para la generación de voces, sitios web como Fakeyou.com permiten crear voces falsas en segundos[21] sin necesidad de conocimiento técnico o avanzado en computación. Es decir, cualquier persona podría crear un audio falso sobre un político, periodista o celebridad para intentar manifestar un mensaje falso de una forma mucho más convincente. Esto tiene usos potenciales en estafas telefónicas, falsificación de identidad, fabricación de

[21] https://fakeyou.com/tts/result/TR:6682zp05ffpq63ww4yeqrcqf19x2f

información falsa, etc. Veritone (2022). Lo importante es entender que en lo que respecta a *Deep Fakes* para la generación de audios existen dos grandes técnicas: la tecnología AD -un ordenador recrea la voz de una persona o personaje específico- o por imitación -un ser humano imitando a otro ser humano, Almutairi & Elgibreen (2022).

Las amenazas de los *Deep Fakes* como "armas" para crear *fake news* altamente convincentes son reales. Hay varios ejemplos conocidos como el de un video de Marc Zuckerberg en el que "admite una conspiración alrededor de los datos personales"[22]; Otro en el que el presidente de Ucrania Zelensky "urge a sus ciudadanos a bajar las armas"[23]. Por ejemplos como estos es que las *Deep Fakes* y las *fake news* son una amenaza creciente para la democracia, Cote (2022).

La propia democratización de la tecnología se posiciona como un elemento acelerador en la generación de contenido, y este hecho incluye de forma indiscutible a las *fake news*. Por tanto los avances tecnológicos y el uso novedoso que pueda darse de las nuevas tecnologías constituyen nuevas oportunidades para que las *fake news* puedan ser creadas y difundidas a velocidades aún mayores.

Los ejemplos anteriores son alarmantes porque son un grupo de contenido que sería completamente imposible hace apenas unos años atrás, con esto no se quiere decir que hubiese sido imposible, claramente con la tecnología de efectos especiales existente hace 10 años sería perfectamente posible generar el mismo resultado que las *Deep Fakes* generan en la actualidad, la única diferencia es el costo, el esfuerzo y el tiempo que toma generar este tipo de contenidos: lo que hace 10 años habría requerido semanas o meses de trabajo, un equipo profesional y altísimos recursos en computación hoy en día solo es necesario de un ordenador casero y del software entrenado adecuadamente para la *Deep Fake* deseada obteniendo un resultado perfecto en tan solo unos minutos. Nunca había sido tan económico fabricar información falsa y es seguro que en el futuro los resultados serán aún mejores, más rápidos y más sofisticados.

[22] https://www.youtube.com/watch?v=cnUd0TpuoXI
[23] https://twitter.com/MikaelThalen/status/1504123674516885507

La automatización y los robots amplifican los efectos de las *fake news*

La automatización es otro de los aspectos tecnológicos con mayor impacto, los robots -perfiles automatizados- en redes sociales han permitido la distribución de *fake news* con fines políticos, lo que hace necesario impedir la proliferación de este tipo de robots con el fin de proteger los procesos democráticos al ser las *fake news* un método de manipulación política, Dias *et al* (2021). Existen además casos históricos como el de la crisis del Golfo en donde se utilizaron distintos bots para la distribución de propaganda y *fake news*, Marc Owen (2019). Lo anterior es posible porque precisamente la tecnología lo permite: es más fácil y barato que nunca para los interesados en distribuir noticias falsas orquestar un "ejército" de robots en internet para que distribuyan toda clase de contenido desinformativo.

Otro caso destacado fue el de la creación de robots y la compra de seguidores para enfrentar protestas laborales de los repartidores, Levy (2022). Nuevamente se encuentra en un caso para el que la tecnología ha sido un facilitador de la desinformación.

Lo anterior hace que el impacto que la tecnología genera a través de las *fake news* es ahora una preocupación política y social en donde se debe considerar que los algoritmos y herramientas más sofisticadas estarán disponibles para todo el público aunque distintas instituciones e individuos intentan limitar su distribución debido a sus posibles peligros, el uso generalizado de la tecnología es inevitable con el paso del tiempo, Chesney & Citron (2019). Esto sustenta además un hecho fundamental: entre más avanzada sea una tecnología, mayor podrá ser la complejidad y escalabilidad del mensaje, a medida que las tecnologías a nivel de software y hardware avancen y sean de acceso público la misma complejidad y efectividad de las *fake news* será mayor: todos están obligados a tener una carrera de armas en el mundo de la información periodística.

Hay que tener en cuenta que esta relación es dinámica -líquida por decirla de otra forma- ya que en sí la tecnología actúa no solo como potenciador de las *fake news*, sino también como un limitador al ofrecer toda clase de recursos para detectarlas y por ende acudir a la acción más recomendada (borrarlas, denegar su distribución, alertar sobre su existencia, etc.) Kozik *et al*. (2022) proponen el uso de tecnologías de *Machine Learning* basadas en NLP

-“Natural Language Processing” o Procesamiento de Lenguaje Natural- como una posible tecnología para la detección de *fake news* de forma automatizada. Este tipo de herramientas e investigaciones son en primera instancia útiles para mitigar la distribución de contenido malintencionado y además sigue sosteniendo la idea ya mencionada: la tecnología altera la forma y el formato de las *fake news* ya que de contar con una tecnología automatizada para detectar noticias falsas inevitablemente éstas tendrán que moldear su redacción, titulares y palabras clave para ser detectadas. Una auténtica lucha de piratas tratando de evadir los filtros de las tecnologías de la información.

Un aspecto interesante es que no existen líneas investigativas que incluyan una reflexión sobre el posible nexo entre las *fake news* y el *Malware* o “software malicioso”-entendiendo por *Malware* cualquier software que tenga una finalidad hostil o intrusiva en cualquier dispositivo-. Esto es esperado ya que actualmente no existe un ejemplo destacado en el que se haya hecho uso de *Malware* para lograr distribuir una *Fake New* pese a la gran relación que existe contemporáneamente entre *fake news* y tecnologías de software como el *Machine Learning*, la inteligencia artificial, los bots, etc.

Uno de los primeros puntos a destacar en este estado de la cuestión es la necesidad de iniciar una reflexión en el mundo académico a través de una prueba de concepto en donde está demostrado que el uso de *Malware* para distribuir *fake news* es perfectamente factible y en donde es solo cuestión de tiempo antes de que este fenómeno empiece a ser visible en nuestra sociedad. Esto abrirá nuevas líneas investigativas además de exponer nuevas necesidades en el campo de la seguridad informática. Un aspecto desarrollado a lo largo de esta tesis doctoral.

Es importante recalcar que la relación entre *fake news* y Software no es nueva. Ya existe evidencia del uso de las *fake news* para facilitar la distribución de *Malware*[24]. También se sabe que existen bots -un tipo de software- para distribuir información falsa, Bradshaw & Howard (2018); y para manipular el discurso político, Shu et al (2020). Otros autores como Kshetri & Voas (2017) sugieren que a través de algoritmos más avanzados será posible crear información falsa perfectamente creíble en un entorno en el que la distribución de las

[24] https://www.trendmicro.com/vinfo/ru/threat-encyclopedia/spam/474/fake-cnn-news-spread-rumors-about-pope

fake news depende más de la ingeniería social -el engaño y la persuasión de las personas- y no tanto de un complejo proceso de ingeniería informática. En otras palabras: los creadores de *fake news* han usado software en el pasado para potenciar los resultados de su contenido desinformativo, y ya hay preocupación sobre el tipo de software que puedan utilizar, el uso del *Malware* es simplemente un paso lógico en este proceso que tomarán tarde o temprano.

Hay casos estudiados como el de Girasa (2020), quien habla sobre el uso de herramientas basadas en Inteligencia Artificial para aumentar la difusión de las *fake news* por parte de gobiernos extranjeros con el fin de intervenir en asuntos políticos; mientras que otros como Nazar & Bustam (2020) entablan la posibilidad de que el software basado en inteligencia artificial pueda ser una herramienta peligrosa ya que existen herramientas Deep Fake como Faceswap -algoritmos de inteligencia artificial que crean contenido falso tal como hacer que cualquier persona aparezca en una película, modificar el discurso de un político, intercambiar rostros de dos personas en una foto, etc. Sample (2020)- que son utilizadas por cientos de personas con resultados realistas y convincentes con la posibilidad latente de que puedan ser utilizados de forma malintencionada.

También existe evidencia sobre la existencia de *Malware* hecho para distribuir información publicitaria a los usuarios; Schultz (2003) habla sobre el "software parasitario", entre ellos el Adware -software que expone publicidad forzosamente en un dispositivo- recalcando el auge de este tipo de software y del generalizado desconocimiento de los usuarios sobre la presencia de este tipo de software en sus ordenadores. Esto expone un hecho fundamental para nuestra prueba de concepto: el *Malware* tiene antecedentes en la distribución de información indeseada, además de lograr persistencia en los dispositivos y "alcance de audiencia" gracias al desconocimiento de los usuarios.

No queda duda de que la tecnología eventualmente encuentra un rol y una función dentro del mundo de las comunicaciones, y esto no excluye a las *fake news*.

A medida que el software disponible consiga mayor complejidad y su barrera de acceso sea menor aumenta la probabilidad de encontrar *Malware* dedicado a la creación o distribución de *fake news*. Por tanto es importante tener conciencia de que el Software será un componente cada vez más utilizado por las Fakes News tanto desde su distribución como

desde su creación. La posibilidad de encontrar troyanos, virus, gusanos o cualquier otro tipo de *Malware* dedicados a distribuir o crear información falsa es cada vez más alta.

Otro aspecto fundamental es la capacidad que tienen las nuevas tecnologías de recabar información personal de toda la población. Nuestra información personal está a merced de las grandes tecnológicas y su capacidad de sacar provecho de estos datos es cada vez mejor. Saber los gustos, intereses, deseos de compra, edad, afinidad política, etc. Son todos aspectos que permiten a los anunciantes -empresas, políticos, entidades gubernamentales, etc.- llegar a audiencias específicas con mensajes específicos.

Cambridge Analytica es el ejemplo perfecto del poder de las audiencias para mejorar el desempeño de las *fake news*. Christopher Wylie -un denunciante del mal actual de Cambridge Analytica- afirmó que la compañía hizo un mal uso de los datos de más de 50 millones de usuarios y que fue precisamente el uso de esta información la que llevó a las *fake news* al "siguiente nivel"[25]. El verdadero escándalo está en la forma en la que Cambridge Analytica extraía los datos de los usuarios: a través de aplicaciones de entretenimiento esta tecnología pedía permisos adicionales de los usuarios en Facebook para extraer entre 4 y 5 mil puntos de datos con el fin de crear un perfil psicológico y de personalidad de cada uno de los usuarios, información que era usada posteriormente para publicitar anuncios electorales, Hu, M. (2020).

Con solo imaginar por un momento la posibilidad de que una empresa pueda saber si una persona siente miedo, incertidumbre, cuál es su ideología política, cuántos amigos tiene, qué forma de pensar tienen sus amigos, etc. y que además con este conocimiento un partido político pueda enviar mensajes personalizados mucho más eficientes según cada perfil psicológico, por ejemplo: "enviar un mensaje sobre el fortalecimiento de las fronteras a quienes estén preocupados por la inmigración ilegal". Esto es un ejemplo clave de cómo la tecnología no solo maximiza el alcance del mensaje, también maximiza su eficacia.

[25] https://content.jwplatform.com/players/xM30iVNe-zffGtjRq.html

Las elecciones estadounidenses del año 2016

Lo expuesto anteriormente denota el aporte que la tecnología tiene inherentemente sobre el desarrollo de las *fake news* y su impacto sobre la sociedad. Esto estuvo reflejado a la perfección en las elecciones presidenciales en Estados Unidos en el año 2016 las cuales son un punto de inflexión en la historia de las *fake news* (Alfonso *et al.* 2019) en donde es posible evidenciar el poder de la información falsa, no solo por su capacidad de manipular, sino por su enorme distribución a través de medios digitales como Google o Meta. Después de todo estos son medios propicios para su distribución, Allcott & Gentzkow (2017). Google y Meta -la compañía propietaria de Facebook- son un duopolio con una audiencia tan grande que los convierte en puntos llamativos para quienes producen *fake news*, ya que estas necesitan distribución con una gran audiencia para cumplir su cometido lo que otorga a las grandes tecnológicas un gran poder y una gran responsabilidad: de repente el mundo entero depende de empresas privadas como Google y Meta -dos de las empresas con una posición dominante en el mercado, Haucap & Heimeshoff (2014)- para evitar la desinformación y la manipulación.

Estas elecciones presidenciales son muy especiales al ser un momento histórico para las *fake news* al ser la cúspide en el que está evidenciado cómo los medios digitales y la tecnología pueden alinearse con la información para moldear las decisiones sociales a escala bajo 3 ideas generales que se pueden extraer del estudio de Grinberg *et al* (2019) en Twitter: 1. Un grupo poblacional con una ideología política específica -derecha y derecha extrema- fue mucho más propenso a compartir *fake news* en redes sociales, 2. El grueso de la población tuvo una exposición mayoritaria a medios de comunicación confiables, y 3. Fueron detectadas cuentas "cyborg" que a diferencia de los bots son más sofisticadas difíciles de detectar que además otorgan un poder de automatización considerable.

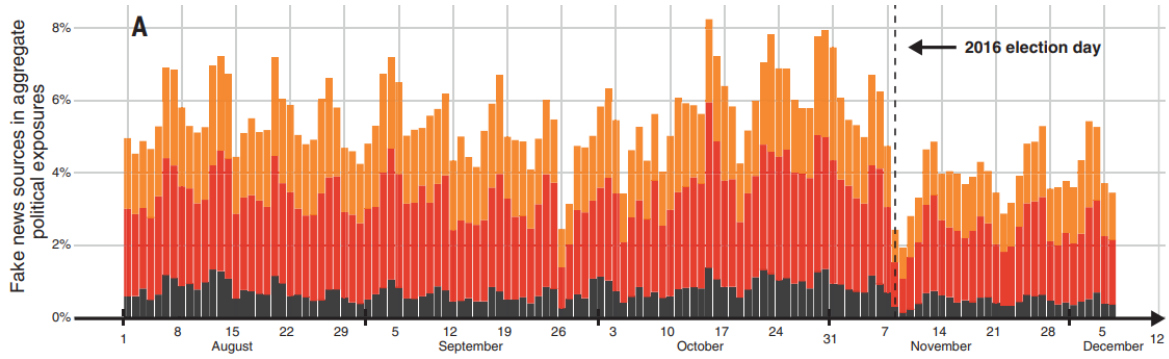
El estudio de Grinberg *et al* muestra algo interesante y es el notorio cambio de exposición que distintos sitios dedicados a la distribución de *fake news* y contenido desinformativo tuvieron un cambio abrupto justo después del día de las elecciones.

Es evidente que una vez terminadas las elecciones los intereses existentes tras la distribución de las *fake news* habían desaparecido: ahora Donald Trump era el nuevo presidente de los Estados Unidos.

Esta correlación no es una prueba explícita de que el partido republicano de los Estados Unidos invirtiera en la distribución de contenido desinformativo, pero sí es un vector que apunta hacia el rol que las *fake news* tuvieron en este proceso electoral.

Figura 5

Exposición de sitios web en Twitter. Incluye sitios de *fake news* conocidos (negro), sitios que distribuyen información falsa por mal proceso editorial (rojo) y sitios sin plena certeza de distribuir información falsa (naranja).



Fuente: Grinberg *et al* (2019)

Es interesante que pese al notorio alcance de las *fake news* existan estudios que exponen otra realidad de los medios digitales: la capacidad de alcanzar audiencias más pequeñas y especializadas, en muchas ocasiones el interés del mensaje no es llegar a una gigantesca audiencia, sino llegar a aquella audiencia que puede convencer y cautivar. Otra interpretación que se extrae del hallazgo de Grinberg *et al* (2019) es que las *fake news* en Twitter no requieren de una gran repetición del mensaje; según este estudio el grueso de la población fue impactada en promedio con 10 URLs durante el último mes de campaña, esto podría significar que con poca exposición y mensajes lo suficientemente impactantes las *fake news* lograron alterar los resultados electorales. No obstante esto es algo imposible de determinar con la información existente. Allcott & Gentzkow (2017) hacen referencia a una idea similar al explicar que existen muchos motivos por los cuales una sola noticia falsa

puede ser mucho más efectiva que un comercial de televisión a la hora de cambiar las decisiones electorales de la población ya que historias como la aprobación del Papa a Donald Trump tendrán más impacto que otros mensajes, aunque medir este impacto de forma precisa no es posible.

Allcott & Gentzkow (2017) también evidencian la importancia de las redes sociales y los buscadores para la distribución de las *fake news* ya que los sitios que distribuyen esta clase de contenidos reciben el 41.8% y el 22% de sus visitantes de estas dos fuentes respectivamente y además en su encuesta sobre la fuente de información más importante para las elecciones de 2016 encontraron que el 13.8% de los encuestados señalaron a las redes sociales y el 14.8% a los sitios web, es decir, casi un tercio de los encuestados admitió que Internet fue su fuente de información predilecta. Es decir, el gran impacto de las *fake news* recae en las nuevas actividades de consumo y en las grandes tecnológicas -resaltando nuevamente la importante relación entre *fake news* y tecnología-.

Es interesante recalcar la importancia de los robots y los métodos automatizados para la distribución de las *fake news* en Twitter. Bovet & Makse (2019) encontraron que aproximadamente el 20% de los tuits publicados por cuentas dedicadas a la distribución de *fake news* eran publicados a través de herramientas no oficiales de Twitter, es decir, mediante procesos automatizados de publicación. Esto solo recalca uno de los aspectos fundamentales para lograr la eficiencia del mensaje: la repetición. Además de esto Bovet & Makse muestran en su estudio que la distribución de *fake news* también fue apoyada por cuentas verificadas como @realDonaldTrump o @PrisonPlanet lo cual juega un papel en el respaldo y credibilidad que estos mensajes reciben por parte de los usuarios.

Lo anterior nos lleva de nuevo hacia la importancia y el poder que tiene la distribución sobre la eficiencia del mensaje: sin robots, automatizaciones y cuentas verificadas dispuestas a distribuir las *fake news* éstas no habrían tenido éxito en las elecciones presidenciales del año 2016. Esta conclusión lógica tiene en todo caso información que puede contradecirla. En su artículo Fox (2019) cita y expone que un estudio de Twitter sugiere que sólo una pequeña fracción de la población de Estados Unidos fue expuesta a las

fake news. Aún siendo cierto, estas investigaciones tienen un problema: asumen que el alcance de las *fake news* se limita a las redes sociales.

En primer lugar, hay que entender que las *fake news* tienen una implicación sobre el voca a voca de las personas, Wisker & McKie (2021). Esto quiere decir que las personas no son un consumidor de contenidos pasivos, sino que dicha información que consumen la comparten con sus personas más cercanas en conversaciones cotidianas e incluso a través de medios de comunicación privados como WhatsApp. Burkhardt (2017) también apunta que las *fake news* son distribuidas por el boca a boca de las personas siendo este un medio de distribución del cual existe poca información. Esto quiere decir que la distribución de las *fake news* en las redes sociales no puede entenderse como un mensaje estático sin posibilidad de llegar más allá de la cuenta de un usuario específico, el poder del mensaje está en justamente mutar y llegar a otras personas a través de distintos medios y métodos.

Lo anterior podría estar reflejado en los resultados del estudio de Gunther *et al* (2018) quienes encontraron que el 25% de las personas que respondieron su encuesta creía que la afirmación "Hillary Clinton está muy mal de salud debido a una enfermedad grave" era cierta y el 35% creía que la afirmación "Durante su tiempo como secretaria de Estado de EE. UU., Hillary Clinton aprobó la venta de armas a los yihadistas islámicos, incluido ISIS" era cierta. Ambas afirmaciones provienen de las *fake news* y por tanto si el impacto de éstas fuera tan pequeño como algunos estudios sugieren sería imposible que encuestas como las practicadas por Gunther *et al* dieran resultados como estos. Las *fake news* obtuvieron una distribución efectiva durante la campaña presidencial del año 2016.

Al ampliar el punto de vista bajo el cual se entiende la distribución de las *fake news* surge como unidad de análisis la "URL", es decir, el artículo que contiene la *Fake New*. Guess (2020) hicieron precisamente un análisis de exposición a estos contenidos y encontraron que en su muestra -personas con 18 años o más en los Estados Unidos- el 44.3% de las personas habían sido expuestas a este tipo de contenidos durante la campaña presidencial.

Gran parte de los estudios que hablan sobre las *fake news* durante las elecciones presidenciales del año 2016 suelen pivotar sobre lo que ha sido mencionado: alta distribución, credibilidad y exposición del contenido. Sin lugar a dudas el éxito de las *fake*

news proviene justamente del rol que la tecnología cumplió a la perfección: potenciar el alcance del mensaje a través de la automatización y las grandes audiencias disponibles en las grandes plataformas tecnológicas.

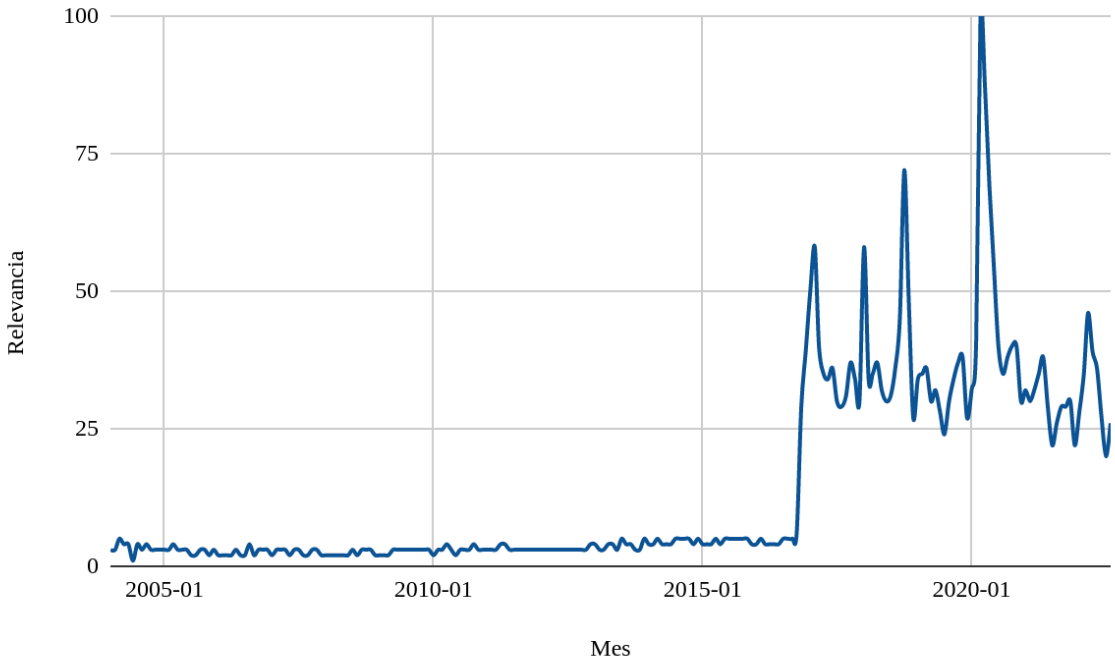
El poder y la influencia de las grandes tecnológicas sobre las campañas presidenciales es notoria; compañías como Microsoft, Google, Facebook y Twitter otorgaron tecnología y servicios publicitarios durante las campañas presidenciales del 2016 con el fin de incrementar sus ingresos y además exponer al mundo la eficiencia de sus servicios, Kreiss & Mcgregor (2017). Claramente las grandes tecnológicas saben que la propaganda política será siempre más efectiva si cuenta con una exposición notoria. En la actualidad "si no estás en Internet, no existes" -dicho popular- y esto lo saben perfectamente. De la misma forma tratar de minimizar cualquier efectividad de las *fake news* gracias a las nuevas tecnologías en este episodio concreto carece de sentido.

Tal poder y capacidad de influencia en las campañas propagandísticas no pasó por desapercibido para el gobierno ruso ya que precisamente durante las elecciones presidenciales del año 2016 la IRA -Agencia Rusa de Investigación de Inteligencia- estuvo asociada a la publicación de distintos mensajes propagandísticos en Facebook con el fin de dividir y polarizar a la población norteamericana y así afectar a las elecciones presidenciales, todo a través del poder de la segmentación basada en datos que Facebook ofrece a sus anunciantes, Ribeiro *et al* (2019).

En lo que respecta a las *fake news* durante las elecciones presidenciales del año 2016 está claro que es desde un punto de vista cuantitativo imposible determinar si la victoria de Donald Trump se debe o no a la distribución de noticias falsas, a las nuevas tecnologías o a la propaganda rusa. Lo importante de estos datos es que su conclusión unánime es que las *fake news* han aumentado su eficacia en el plano político y social gracias a 3 aspectos fundamentales: el fácil acceso a grandes volúmenes de datos, las gigantescas audiencias disponibles para cualquier persona o entidad, la fuerte automatización a través de robots y la enorme personalización del mensaje gracias al uso de datos personales. Las *fake news* tienen muchas más herramientas para cumplir con sus objetivos y el año 2016 fue un capítulo clave para demostrarlo.

Figura 6

Relevancia de la palabra clave "fake news"



Fuente: datos de trends.google.com para la palabra clave "fake news" a nivel global, gráfico de elaboración propia.

Otro aspecto fundamental está en los datos disponibles ya que aunque el uso de la palabra clave *fake news* es muy antiguo en la literatura científica, su uso en el buscador de Google tiene un incremento notorio a partir de diciembre de 2016. En otras palabras: ha sido tras las las elecciones presidenciales de los Estados Unidos en el año 2016 el momento clave en

el que se inicia el uso frecuente de este término ya que no fue sino hasta este punto de la historia de la comunicación en el que se da un uso de las *fake news* a una escala tan grande, lo que evidencia el creciente problema existente en la distribución digital de información que no puede ser verificada.

Figura 7

Temas y consultas relacionadas con la palabra clave "fake news" a nivel global desde el año 2004

Temas relacionados	Principales		Consultas relacionadas	En aumento
1 Noticia - Formato de radio	100		1 trump news	Aumento puntual
2 Fake news - Tema	88		2 fake news trump	Aumento puntual
3 Archivo falso - Tema	33		3 trump	Aumento puntual
4 Donald Trump - Presidente 45.º de los Estados...	5		4 fake news media	Aumento puntual
5 Medio de comunicación - Tema	3		5 facebook fake news	Aumento puntual

Fuente: datos y tablas generadas con datos de trends.google.com. Datos desde el año 2004 hasta la actualidad. Tablas consultadas el día 4 de septiembre del 2022.

La detección de las *fake news*

Las investigaciones científicas enfocadas en la detección de las *fake news* son populares. Durante este trabajo de investigación se han encontrado numerosos estudios y propuestas centradas en mecanismos que utilizan *Machine Learning*, Inteligencia Artificial, NLP, análisis de lenguaje, etc. con el fin de detectar si un contenido específico es una *Fake New*.

Un buen ejemplo es el de Li *et al* (2021) quienes consiguieron diseñar un modelo de Deep Learning de auto-aprendizaje y semi-supervisado con el cual consiguieron una precisión notoriamente alta. Es importante mencionar que este modelo solo está diseñado para detectar *fake news* en redes sociales y además fue entrenado con un set de datos ya existente, es decir, sus resultados no fueron probados en un campo real aplicable sino en un entorno controlado. Eso denota de por sí una de las grandes barreras que el mundo académico enfrenta al momento de investigar la detección de las *fake news*: no existe un

acceso sencillo al entorno cerrado de Facebook o Twitter, la información disponible siempre será una base de datos preestablecida, lo cual limita de forma significativa el avance de estas técnicas.

Aún así otros investigadores apuntan a otros hechos evidentes que son útiles en el proceso de detección, Shrestha & Spezzano (2021) mencionan distintos factores comunes que las *fake news* suelen compartir: suelen tener menos contenido, menos citas y menor puntuación, requieren de un menor nivel de educación para poder leerlas y comprenderlas, sus títulos suelen ser largos con palabras cortas y usando mayúsculas siendo éstos los que diferencian con mayor facilidad a una noticia real de una falsa. Ciertamente estos hallazgos nos dan a entender que el análisis semántico y lingüístico del contenido es la primera y más útil herramienta para detectar el contenido desinformativo, pero hay un problema inherente con esto: estas técnicas serán útiles solo si la forma en la que se redactan *fake news* no cambia en el futuro; ciertamente no existe garantía alguna de que su desarrollo y redacción no vaya a ser más sofisticado en el futuro. Ciertamente un contenido bien escrito no es garantía de que sea verídico.

La detección de *fake news* además de ser un reto técnico pasa por ser un reto de arquitectura tecnológica y autores como Kozik *et al* (2022) lo ejemplifican en su trabajo investigativo al proponer una estructura de tecnologías integradas entre sí para lograr la detección de *fake news*, parte de esto según explican recae sobre el comportamiento de los modelos de aprendizaje ya que un único gran modelo funciona de forma excepcional en un único dominio, pero hacerlo trabajar en otros es una tarea tediosa y compleja. Esto hace que tenga sentido su propuesta estructural ante el problema de la detección de las *fake news*.

Otros investigadores como Shahid *et al* (2022) han explorado la posibilidad de detectar las cuentas robot encargadas de distribuir las *fake news* a través de las redes sociales a través del análisis de su personalidad, su información histórica y su credibilidad. No obstante como bien mencionan es un campo investigativo poco explorado y difícil de desarrollar ya que las cuentas robot encargadas de distribuir las *fake news* pueden mimetizar fácilmente el comportamiento humano.

También se debe considerar el aspecto psicológico detrás del texto ya que después de todo las *fake news* suelen apuntar a las emociones, Shrestha & Spezzano (2021), con lo cual

hacer un procesamiento de lenguaje natural basado en las emociones que transmite un texto también puede ser un factor a tener en cuenta en la detección de las noticias falsas.

Lo que es posible observar es que en general las investigaciones relacionadas con la detección de las *fake news* afrontan los mismos retos: la homogeneidad de la información -cada autor de *fake news* tiene un estilo diferente-, la dificultad para acceder a datos mucho más extensos, la imposibilidad de aplicar los modelos en un entorno real y la gran divergencia tecnológica en donde existen muchas soluciones para un mismo problema, pero ninguna solución expone un método que funcione de forma universal. Esto implica que la detección de las *fake news* deberá ser en el futuro un trabajo especializado que no puede depender -de momento- al 100% de las tecnologías automatizadas.

Llama la atención que pocas investigaciones se enfoquen en la reputación de las fuentes que publican información, después de todo la confianza y la reputación de un sitio web concreto es un factor considerable que ayuda a detectar las *fake news*. En este caso hay investigaciones como las de Xu *et al* (2020) quienes usando datos del top de sitios web de Alexa encontraron que los sitios web dedicados a la publicación de *fake news* tienen comportamientos en común como la despublicación de contenido y el utilizar dominios web mucho más jóvenes. Esto simplemente confirma que la detección de las *fake news* no se limita al análisis de contenido en sí mismo, sino también al análisis de sus fuentes, autores y sitios web.

La detección de las *fake news* no se limita a un análisis textual, también existen investigaciones que exploran la posibilidad de detectar imágenes falsas; tal como refleja la relación entre *fake news* y tecnología la creciente capacidad de falsificar fotos, videos y audios mediante técnicas como el Deep Fake también son formas en las que una *Fake New* puede dar comienzo. Masciari *et al* (2020) exploran dicha probabilidad y proponen un esquema mediante el cual es posible detectar imágenes falsas utilizando Inteligencia Artificial usando procesadores de lenguaje natural y multimedia. Ciertamente estos procesos de detección son mucho más complejos y difíciles de desarrollar que un modelo basado en textos y sin duda es uno de los campos que más desarrollo necesita en este campo de estudio.

En el caso de las *Deep Fakes* hay dificultades distintas en especial en el caso de los videos ya que estos son mucho más difíciles de detectar. A pesar de ello existen a nivel general dos métodos tal como mencionan Pu *et al* (2022): detección a nivel de "todo el video" y la detección de fotogramas específicos; en todo estos autores exponen que ninguno de los dos métodos son una solución universal y que tienden a tener dificultades en detectar *Deep Fakes* en contextos específicos. Un ejemplo de ello -explicado también en el artículo de Pu *et al*- es la detección de píxeles modificados en cada uno de los fotogramas del vídeo, aunque este tipo de métodos han sido superados por las técnicas de *Deep Fakes* modernas, Jung *et al* (2020).

Existen otros métodos como la detección de patrones en el parpadeo de las personas en cada vídeo, un método novedoso que compara la forma en la que una Inteligencia Artificial intenta mimetizar el parpadeo humano, el cual tiene un comportamiento específico según el contexto y cada ser humano en sí; pese a ello este método también tiene retos ya que este método no sería aplicable si en el video hay una persona con enfermedades mentales o problemas en el sistema nervioso, Jung *et al* (2020).

En lo que respecta a la detección de *Deep Fakes* auditivos los métodos de detección se basan también en Inteligencia Artificial, *Deep Learning* y *Machine Learning*, dichos métodos son muy diversos, pero todos recaen en las mismas tecnologías teniendo en común en muchos casos la dificultad de tener que procesar grandes volúmenes de datos involucrando en muchos casos un trabajo manual, Almutairi & Elgibreen (2022). En todo caso los autores también mencionan la necesidad existente de mejorar y desarrollar aún más este campo de investigación de la detección de *Deep Fakes*.

Lo que se extrae de estos autores es que los métodos de detección de *fake news* en la actualidad se basan -por lo general- en modelos estadísticos o matemáticos basados en Inteligencia Artificial, *Machine Learning* o Deep Learning para lograr procesar a escala contenido visual, textual, auditivo o audiovisual con el fin de detectar información imprecisa -en el caso del contenido textual- o información modificada -para el caso de los contenidos auditivos, visuales o audiovisuales-. Dicho esto, es importante recalcar que los métodos actuales distan mucho de poder aplicarse de forma exitosa en la práctica, es decir, son métodos que aún se encuentran en una fase meramente teórica con pruebas muy

limitadas al probarse con sets de datos bastante específicos. No solo eso, no existen métodos fiables para detectar *fake news* en formato de video o de audio -cosa muy diferente de un *Deep Fake*-, por ejemplo si hubiese un video viral en WhatsApp que habla sobre "el daño al ADN que hace la vacuna contra el COVID-19" no sería posible clasificarlo de forma automatizada como una *Fake New* usando alguno de los métodos citados anteriormente.

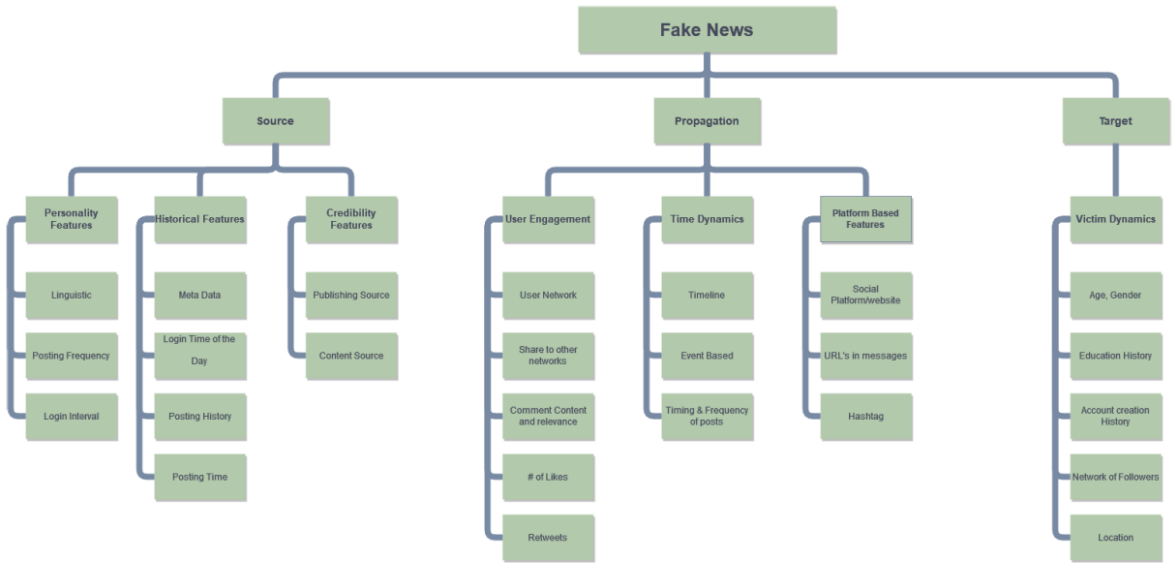
Finalmente la detección de las *fake news* es un campo de estudio cuyo desarrollo es constante, las nuevas tecnologías son cada vez más sofisticadas y los métodos de engaño, persuasión y redacción cambian con el tiempo, por tanto las técnicas actuales de detección podrían no funcionar en el futuro -caso que ya ha ocurrido con las *Deep Fakes*-, esto implica que el reto más grande de los métodos actuales no es su precisión en la detección de *fake news* actuales, sino su resiliencia en el tiempo y la probabilidad que tienen de adaptarse a nuevas formas de contenido.

También se han encontrado otros autores con conclusiones similares, por ejemplo Couto & Modesto (2020) encontraron una correlación positiva entre el uso de Facebook y el activismo y la radicalización política, además que por la mismo comportamiento del *News Feed* de Facebook -facilitar la existencia de cámaras de eco- la constante interacción en distintas publicaciones políticas de Facebook son lo que llevan a la radicalización puesto que este entorno de interacción carece de referencias externas u opiniones disidentes.

La detección de las *fake news* también pasa por el análisis de las cuentas dedicadas a distribuir dicho contenido. En este sentido hay muchas investigaciones relacionadas que se dedican a esclarecer y categorizar todos los aspectos asociados a dichas cuentas. El caso de Shahid *et al* es un ejemplo destacado.

Figura 8

Taxonomía sobre las características de la difusión de noticias falsas basada en las cuentas de los difusores de noticias falsas



Fuente: Shahid et al (2022)

En este caso la figura 8 muestra una taxonomía completa de las características propias de una difusión de las *fake news* teniendo en cuenta precisamente las características de los actores difusores que existen en la red. Es interesante recalcar la existencia de elementos demográficos propios de una audiencia -edad, género, educación, ubicación, etc.- ya que claramente las *fake news* siendo un material comunicativo deben de tener en determinado momento un público objetivo, o dicho de otra forma: la forma y el contenido de una *Fake New* depende de quién la va a leer. Claramente estos son factores que involucran la forma

en la que deben de ser detectadas y además son elementos visibles en las cuentas dedicadas a su propagación.

El *News Feed* de Facebook

Facebook se ha posicionado como uno de los principales distribuidores de noticias a nivel mundial al ser un método fácil y accesible para leer el contenido más reciente a nivel local y mundial -como se ha visto gracias a varias investigaciones presentes en este marco teórico las redes sociales son una fuente de información habitual y Facebook es sencillamente una de las más grandes-. Esto hace que de manera implícita muchos medios de comunicación dependan de este -y otros algoritmos- para obtener exposición. Si el algoritmo no considera que tu contenido no sea importante, no serás visible en Facebook. Quien no se adapte al algoritmo, desaparecerá de Facebook, lo que es una manipulación mediante la invisibilización.

Las investigaciones académicas vigentes alrededor de los algoritmos del *News Feed* de Facebook ya demuestran la dependencia -directa o indirecta- que tienen los medios de tener visibilidad en Facebook. Un ejemplo de ello es la demostración de que los medios de comunicación en los Estados Unidos aumentaron la cantidad de videos nativos como respuesta a los cambios de los algoritmos que dieron prioridad a este formato dentro de Facebook, Tandoc & Maitra (2018). Que los usuarios disfrutan del *News Feed* pese a los inconvenientes en políticas y prácticas de privacidad, Hoadley *et al* (2010). Esto evidencia el impacto y relevancia que tiene *News Feed*.

News Feed, una aplicación fundamental del Facebook actual existe desde 2006 y es considerado como revolucionario al ser el primer "feed [proveedor]" de contenido social en la historia (D'Onfro, 2016). Su valor reside en su capacidad para distribuir los contenidos disponibles en un "feed [proveedor]" central y personalizarlo para cada usuario. Jamás se había logrado una herramienta así. Antes de *News Feed* el consumo de contenidos en Facebook (y otras redes sociales) se basaba en una navegación por perfiles para encontrar las publicaciones que cada usuario deseaba ver. En ese esquema el alcance del contenido estaba limitado por la capacidad de los usuarios para encontrarlo: a través de otros perfiles sociales o desde fuentes externas (buscadores, feeds RSS, foros, etc.). El hito de *News Feed*

es que un usuario puede acceder a contenidos de páginas y usuarios sin la necesidad de acceder a sus perfiles directamente, de ahí su enorme capacidad como influencer. ¿Por qué? Porque todos los contenidos se juntan en un único proveedor -feed-, lo que supone un cambio radical en el esquema de distribución de contenidos online.

El impacto de la estrategia *News Feed* es tal que inspiró el camino de otras redes como Twitter, Pinterest e Instagram (Manjoo, 2013), demostrando su influencia no sólo en los medios digitales sino en las masas al tener un papel fundamental en que las personas permanezcan más tiempo enganchadas en las redes sociales. La influyente revista Forbes alertaba en 2018 que tras unos cambios en el *News Feed* la interacción de los usuarios se redujo a la mitad: "Facebook Engagement Sharply Drops 50% Over Last 18 Months (Erskine, 2018)". Esto demuestra de forma explícita la importancia de esta herramienta sobre la permanencia en Facebook, y por ende su valor en su modelo de negocio.

Dado que el *News Feed* es la principal herramienta de distribución de contenidos en Facebook es evidente entender que su relación con el éxito de las *fake news* es perfectamente posible más aún en un contexto de consumo de contenido donde la poca profundidad y una inmensa variedad de material promueve el consumo de noticias "simplemente leyendo su titular". Geeng *et al* (2020) resaltan en su investigación este fenómeno en el que encontraron que un número significativo de participantes que consumieron *fake news* en Facebook no profundizaron sobre el contenido consumido.

David *et al* (2019) encontraron en su estudio -de una audiencia filipina- una asociación positiva entre la exposición de noticias políticas en Facebook con la participación activa en temas de este tópico. Es cierto que un estudio de este estilo no puede extenderse a todos los mercados y culturas y que en un entorno con miles o millones de factores que pueden afectar a este tipo de estudios es muy difícil llegar a una conclusión precisa, termina siendo interesante hallar dicha correlación. Después de todo, los ojos del mundo entero no estarían en Facebook si su distribución de contenidos no tuviera una influencia política.

Además de esto cada usuario de Facebook está influido fuertemente por sus amigos dentro de esta red social frente a los contenidos que consume; según a qué contenidos sus amigos

interactúan compartiendo, comentando o haciendo un "like" una persona se verá fuertemente influenciada a consumir determinados contenidos, Anspach (2017).

Está claro que el funcionamiento actual del *News Feed* más que ser un método llamativo para la distribución de *fake news* es además un entorno favorable para aumentar su credibilidad ya que propicia un entorno de vulnerabilidad para el usuario: pocas opiniones disidentes, escasas referencias o validación de veracidad de contenido y además un algoritmo que siempre le expondrá a "más de lo mismo", es decir, el entorno perfecto para repetir un mensaje y convencerle de algo.

Entre toda la información conocida se sabe que el caso de las *fake news* afectó al *News Feed* de Facebook a tal punto que la compañía ha invertido esfuerzos en combatir la diseminación de noticias falsas a través de este medio[26].

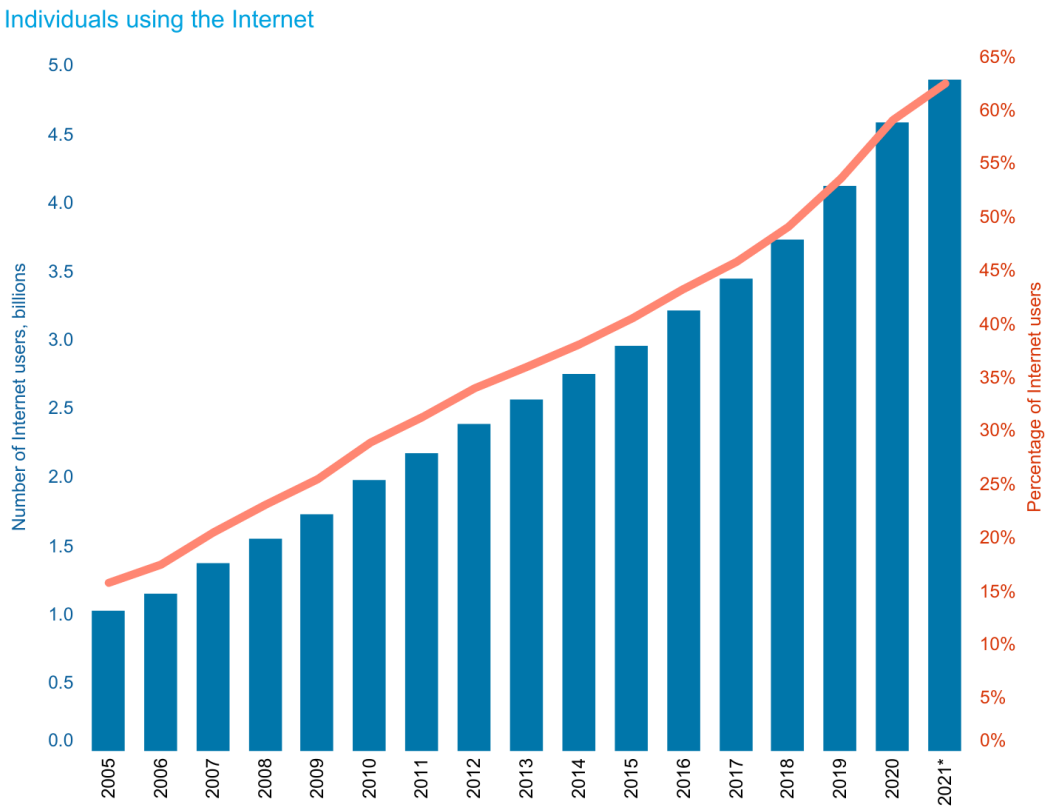
Periodismo y nuevos modelos de consumo de contenido

La distribución de las *fake news* a través de los medios digitales tiene una asociación directa con los nuevos métodos de consumo que las personas presentan hoy en día, es decir, la forma en la que el grueso de la población consume noticias y material periodístico es completamente diferente.

[26] https://www.proquest.com/docview/1863207571?pq-origsite=primo

Figura 9

Evolución anual en miles de millones de usuarios con acceso a internet estimado por la Unión Internacional de Telecomunicaciones.

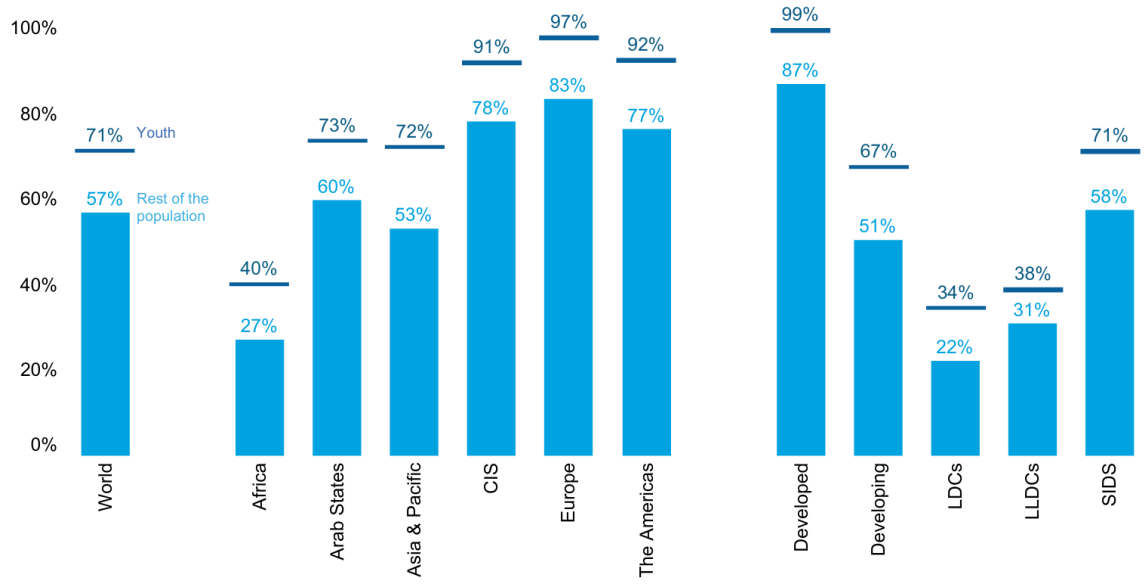


Fuente y elaboración: International Telecommunication Union. (2021)

Como ya muestran los datos de la UIT -Unión Internacional de Telecomunicaciones- en la Figura 9 el acceso a internet es una constante en la población global, la penetración de Internet como herramienta de consumo de contenidos ya supera el 50% de la población global y su crecimiento no ha hecho más que acelerarse tras la pandemia del año 2020.

Figura 10

Porcentaje de individuos usando Internet por región dividido por grupos de edad: jóvenes (15 a 24 años de edad) y el resto de la población (menos de 15 o más de 24 años de edad) en el año 2020.



Fuente y elaboración: International Telecommunication Union. (2021)

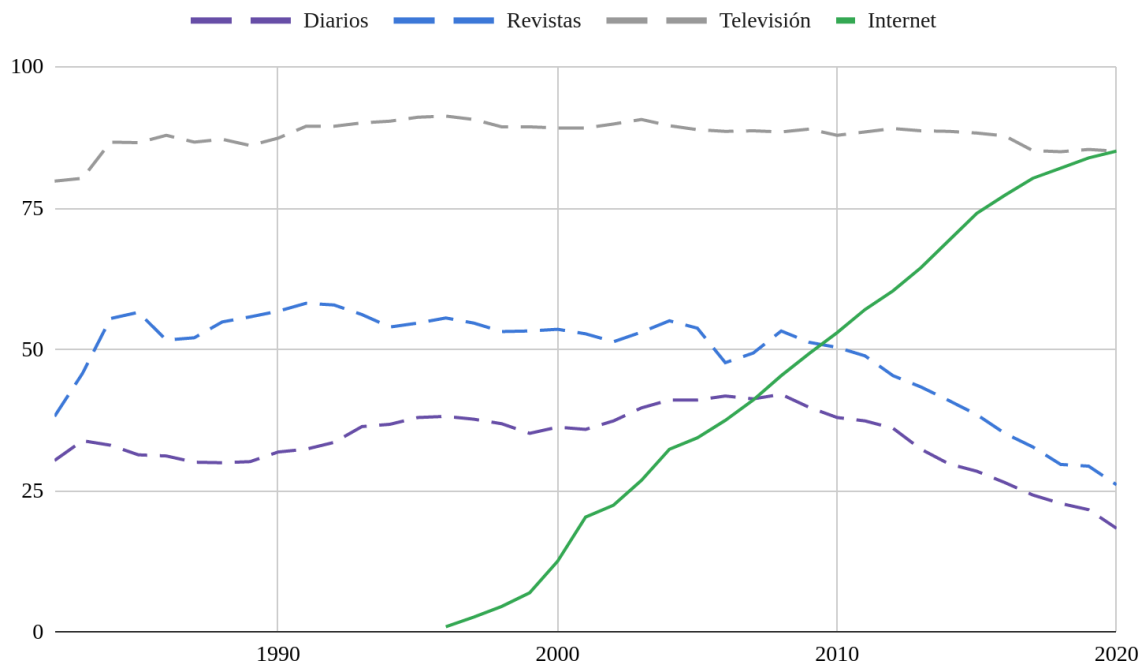
Dicho uso de Internet tiene diferencias sustanciales entre regiones geográficas y grupos de edad ya que evidentemente las regiones con mayor desarrollo económico -Developed- presentan una mayor penetración de Internet -87%-, pero además de ello los jóvenes casi en su totalidad presentan un consumo de internet -99%-. Esto hace que la distribución de las *fake news* a través de medios digitales tenga un mayor impacto precisamente en los países desarrollados.

Estas estadísticas en sí mismas son significativas para el periodismo porque el aumento generalizado del uso de Internet ha supuesto también una transformación en todos los medios, especialmente para los diarios y las revistas -medios impresos-. Así lo refleja los

datos de la AIMC para España en la Figura 10 en donde el crecimiento de Internet ha supuesto una caída constante -que parece continuará en el futuro- para estos medios.

Figura 11

Evolución de la audiencia general de medios en España por porcentaje de penetración.



Fuente: AIMC (2021), gráfico de elaboración propia

Por supuesto un mayor consumo de Internet supone un mayor consumo de periódicos en medios digitales, sin embargo las redes sociales han actuado como una barrera ya que la mayoría de usuarios se han conformado con la información recibida en sus redes sociales -Facebook y Twitter- reduciendo la cantidad de lectores totales; esto trae retos cada vez más crecientes para los periódicos y periodistas ya que el consumo de contenidos en Internet exige la aplicación de nuevas soluciones para rentabilizar su actividad, Marcos Recio *et al* (2018).

Los cambios tecnológicos también traen consigo un cambio en los métodos de consumo tal como Ravettino Destefanis (2019) menciona en su artículo ya que el contenido hipertextual trae una forma de leer que promueve la flexibilidad y la interacción dejando a un lado la permanencia en el contenido. Esto tiene perfecto sentido ya que con la llamada economía de la atención y la creciente abundancia de contenidos es imposible para cualquier persona hacer una lectura concienzuda de cualquier artículo noticioso; de hecho, varios artículos y estudios señalan que los usuarios aumentan el consumo de contenidos en redes sociales mientras que limitan la lectura al titular[27].

Para el periodismo está claro que los entornos digitales de consumo de contenidos no trajeron una cantidad sustancial de nuevos lectores dedicados, más bien ha puesto a la disciplina en un entorno de alta competitividad en donde la atención -ya bastante limitada por parte de los usuarios- lo es todo.

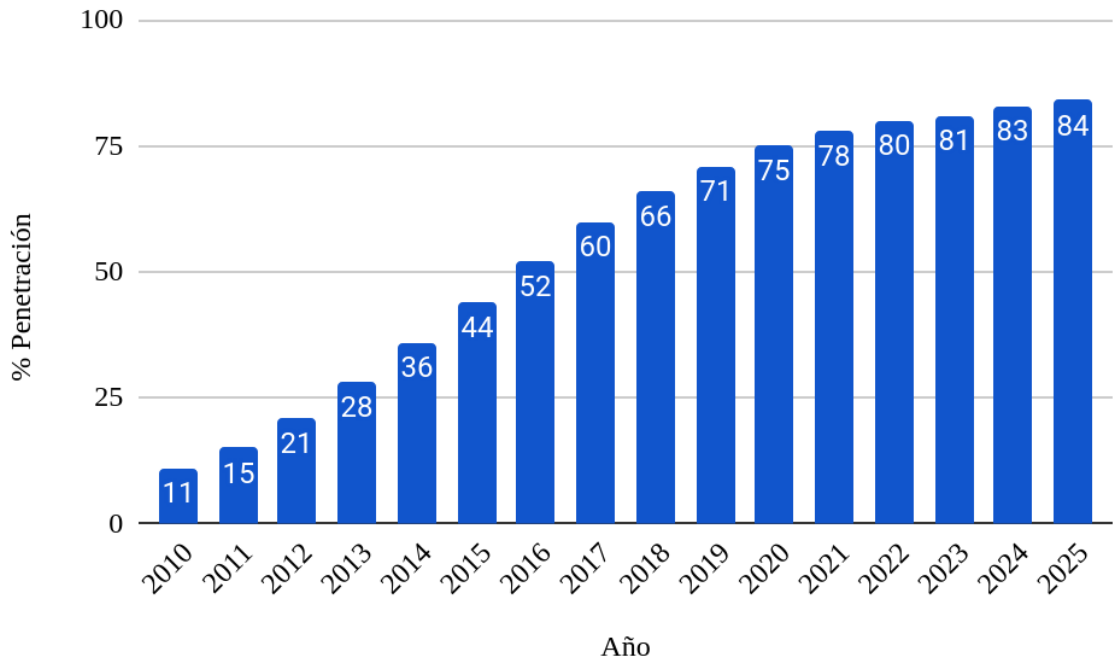
El impacto de la tecnología sobre el periodismo también involucra al mismo dispositivo sobre el cual el usuario hace el consumo de contenidos, después de todo no es lo mismo leer un artículo o explorar el contenido de un periódico usando un ordenador de escritorio a un teléfono móvil. Para el caso de los periódicos entre más clics requiera un usuario para completar una tarea más compleja le parecerá, incluso si un usuario requiere de muchos pasos para leer un contenido optarán por usar un buscado web, Jiménez Iglesias *et al* (2018).

Lo anterior implica que los periódicos tienen una ardua tarea al momento de captar usuarios y retenerlos en su sitio, deben asumir que la gran mayoría de sus visitantes solo quieren acceder a la información de una forma inmediata. Esto también es un punto que facilita el trabajo de las *fake news*, después de todo si la lectura es corta y en muchas ocasiones limitada al titular de la noticia entonces estamos en un escenario en el que los titulares llamativos, amarillistas y potencialmente engañosos pueden obtener mayor atención del usuario y con ello mayores interacciones en redes sociales. Estamos en un escenario que no favorece al periodismo de calidad y profundidad, es un mercado que favorece la cantidad sobre la calidad.

[27] https://www.genbeta.com/redes-sociales-y-comunidades/el-56-8-de-los-lectores-espanoles-de-prensa-se-informa-a-traves-de-redes-sociales-aunque-solo-lee-titulares-y-alguna-noticia

Figura 12

Estimación evolutiva de la penetración de los dispositivos móviles inteligentes en Europa.



Fuente: Statista. (2021), gráfico de elaboración propia

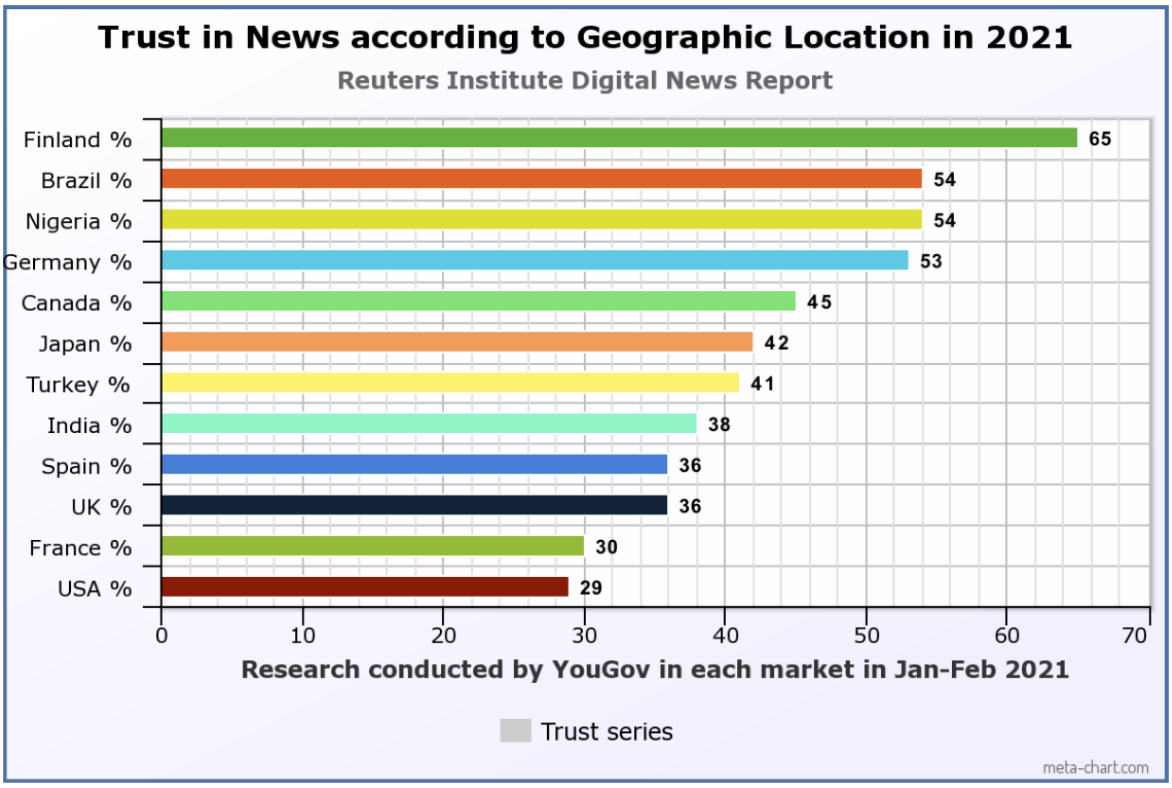
Un aspecto interesante es que la disminución de lectura impresa en revistas y periódicos no inicia inmediatamente después del inicio de la penetración del Internet, sino que ocurre a partir del año 2010, año en el que el crecimiento vertiginoso de los *Smartphones* o teléfonos inteligentes tiene lugar en Europa -Figura 12-. Por lo que el periodismo no se ha enfrentado en la última década a un único cambio, sino a una serie de transformaciones tecnológicas que han influido notoriamente en los hábitos de consumo de gran parte de la población.

Fake news y sociedad

Ciertamente el contexto social bajo el cual una *fake news* es difundida cambia su eficiencia y así mismo el interés de distintos actores en invertir en ellas. Si una sociedad tiene una fuerte confianza en sus fuentes de información oficiales entonces las *fake news* tendrán un menor porcentaje de éxito.

Figura 13

Confianza en las noticias según la ubicación geográfica en 2021.



Fuente: Shahid et al (2022).

Como se observa en en la Figura 13, Estados Unidos figura como el país con menor confianza en sus fuentes noticiosas lo cual no supone una sorpresa el hecho de que sea precisamente en este país en donde más estudios y efectos de las *fake news* son evidenciados en la actualidad. Es importante considerar que estos indicadores pueden tener

una correlación pero no necesariamente una regla que rige con exactitud el comportamiento de las noticias falsas. Por ejemplo, Brasil, el segundo país en esta gráfica, sigue siendo una ubicación en la que 4 de cada 10 personas afirman recibir noticias falsas[28]. Lo destacado de esta información es principalmente la amplia diferencia en el comportamiento social que tiene cada país de forma individual y por tanto cualquier análisis global sobre las *fake news* debería considerar esto.

Conceptos

- **NLP:** Una rama de las ciencias de la computación que combina modelos estadísticos, de *Machine Learning* y Deep Learning con el fin de procesar el lenguaje humano -en formato de texto o de voz- con el fin de que una computadora pueda "entender" el significado de un mensaje a nivel de intención y sentimiento. IBM (2020)
- ***Machine Learning*:** Es una rama de la Inteligencia Artificial que a partir de distintos datos y algoritmos mimetiza el aprendizaje humano para que con el tiempo consiga incrementar la precisión de sus resultados. IBM (2020)
- **FoMO:** Fear of Missing Out, se define como una aprehensión de que otros puedan tener experiencias agradables y no estar presente en ellas, lo cual lleva al deseo de estar conectado constantemente con lo que otros están haciendo. Przybylski *et al* (2013)
- **SMF:** Social Media Fatigue, la tendencia de un usuario de alejarse de las redes sociales y reducir su participación en ellas cuando se siente saturado de información. Bright *et al* (2015)
- **Malware:** Cualquier software que tenga una finalidad hostil o intrusiva en cualquier dispositivo. Oracle (2022).
- **Adware:** Software que expone publicidad forzosamente en un dispositivo, Kaspersky (2022).
- **Cyborg:** Perfil en redes sociales gestionado parcialmente por un robot y un ser humano, Shahid *et al* (2022).

[28] https://www.cnnbrasil.com.br/nacional/4-em-cada-10-brasileiros-afirmam-receber-fake-news-diariamente/

METODOLOGÍA

Esta tesis está dividida en distintos capítulos cuya información abarca un aspecto específico de las dos grandes gigantes tecnológicas más conocidas en la actualidad: Google y Meta. Por este motivo cada capítulo carga con un mecanismo metodológico específico que estará detallado dentro del mismo. En términos generales las metodologías usadas en esta investigación combinan aspectos cualitativos -como el análisis textual o la interpretación de resultados- y cuantitativa -extracción de datos numéricos de APIs-.

A nivel cualitativo existe un trabajo de recolección y organización de información de distintas fuentes, dicho trabajo que es en última instancia una documentación ordenada de una información ya existente puede observarse en el capítulo "Los algoritmos del *News Feed* de Facebook" en el cual se documentan los cambios más importantes alrededor del *News Feed* y sus distintas actualizaciones con mayor impacto a nivel histórico. También existe una aplicación similar en el capítulo sobre "Google y las *fake news*" al tener por primera vez de forma ordenada y analizada las distintas tecnologías y metodologías que Google podría estar usando en su buscador.

Lo importante de este análisis cualitativo está no solo en la recolección de documentación que hasta la fecha se encontraba repartida en distintas fuentes, sino también en su orden y valoración según su impacto en el contexto de las *fake news*. Cada capítulo especifica cómo se organiza y centraliza la información aunque en términos generales se busca tener una organización cronológica de los datos con una fuente académica o profesional que le respalde. Para los casos en los que exista más de una fuente haciendo referencia a un tema concreto, se optó por tomar la fuente más relevante con el fin de lograr un nivel de síntesis apropiado.

Finalmente tras documentar y ordenar la información disponible cada capítulo cuenta con una valoración y un análisis de dicha información, dicho análisis conecta cada uno de los aspectos hallados para entregar conclusiones importantes en el contexto de la desinformación y su distribución a través de los medios digitales.

A nivel cuantitativo existen dos metodologías, la primera de ellas fue la extracción de datos a través de APIs los cuales fueron organizados en grandes bases de datos que, a través de

herramientas de inteligencia de datos, permiten estudiar fenómenos concretos. Por ejemplo en el capítulo de "Los algoritmos del *News Feed* de Facebook" fueron descargados años de publicaciones de los periódicos más destacados de España, información que fue ordenada, graficada y analizada para poder contrastar con cada uno de los cambios algorítmicos de Facebook. Este ejemplo concreto trae grandes ventajas ya que los datos disponibles no son de un tercero, son del mismísimo Facebook, lo cual da aún más confianza sobre su precisión y uso.

Para el caso del capítulo "Google y el filtro de las *fake news*" fueron descargados datos de cientos de dominios desde varias fuentes de datos altamente confiables -Zvelo, Ahrefs y Sistrix- datos que fueron procesados y normalizados a través de distintas técnicas de limpieza de datos y fueron unificados en un gran set de datos. Este proceso de limpieza supone un nivel de complejidad alto que a su vez trae una visión única del comportamiento de una serie de sitios web que en su resultado final permite esclarecer cómo Google abordó en un momento puntual el posicionamiento de las *fake news* en su buscador.

Aquí se vuelve a resaltar lo ya mencionado, metodológicamente se combina una estricta recolección de datos en un método completamente cuantitativo y orientado a una analítica meramente matemática que es contrastado con una documentación ordenada y analizada cualitativamente. La combinación de ambos análisis -cualitativo y cuantitativo- es lo que trae a flote varias de las conclusiones de cada uno de los capítulos.

La última metodología es cuantitativa y contiene en sí misma un enfoque experimental y está aplicada en el capítulo de "*Malware* y *fake news*". Esta metodología consiste en la programación de un software -extención de Google Chrome- que es publicada en la *Chrome Web Store* con el fin de que sea aprobada. La parte experimental está en que dicho software estará diseñado para insertar información falsa en lavanguardia.com y el objetivo es observar hasta qué punto Google puede detectar un software malicioso con estas características.

LOS ALGORITMOS DEL *NEWS FEED* DE FACEBOOK. ANÁLISIS DE SU INFLUENCIA MEDIÁTICA EN LA ERA DE LAS *FAKE NEWS*

Introducción

En este capítulo abordamos un análisis en profundidad sobre el *News Feed* de Facebook, servicio que está posicionado como la mayor red social del mundo con una capacidad de permea la comunicación entre usuarios para lograr influenciar la forma en que los medios se comunican. Es el gran influencer y así lo evidencian los datos analizados.

Este análisis está compuesto por dos partes: 1. La primera cartografía muestra cada uno de los cambios del algoritmo de Facebook relacionado con su *News Feed*. 2. Un análisis que busca demostrar que estos cambios realmente influyen en los medios. En este proceso recolectamos 122.869 publicaciones, mediante el Graph API de Facebook, en un periodo que va desde 2008 a 2019. Seleccionamos los 20 periódicos más importantes según el EGM de 2008. Los resultados son concluyentes: los periodistas detectan cambios del algoritmo y modifican sus contenidos para tener más interacciones y sobre todo influencia. El estudio se centra en el *News Feed*, pero rastrea cómo esos cambios en el algoritmo pueden estar detrás del aumento de las *fake news*.

Nunca un medio había sido tan gigantesco: en abril de 2019 Facebook, según su web, tenía casi 2.400 millones de usuarios (casi 3.000 millones según otras fuentes[29]). Ni tampoco tan poderoso: posee una influencia líquida que trasciende fronteras políticas, ideológicas o físicas. Probablemente es el "*News Feed*" (el alimentador de contenidos de Facebook) uno de los principales motivos que hacen de esta red social un fenómeno tan importante en la actualidad. Son los algoritmos que seleccionan y filtran la información distribuida los principales responsables de la posición privilegiada en la que se encuentra Facebook en la actualidad. Esto hace que su investigación y análisis sea tan llamativo.

Como está plasmado en el estado de la cuestión la relación entre Facebook y los medios de comunicación es una relación de dependencia: medios como los periódicos necesitan visibilidad en Facebook -y en otras redes sociales- para obtener ingresos al ser una

[29] https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/

importante fuente de clics y tráfico web. Así mismo la relación entre usuarios y Facebook no es una relación "saludable" al tener serias implicaciones en términos de privacidad[30] y adicción[31]. Esto hace especialmente llamativo este capítulo de la tesis doctoral la que se desarrollará un escrutinio muy detallado cómo el *News Feed* de Facebook tiene un papel fundamental en la distribución del contenido y qué tipos de cambios han existido a lo largo de su historia.

En este capítulo se describe la cronología evolutiva de los algoritmos del "*News Feed*" para comprender cómo el poder de influencia de Facebook -junto a Google- logra controlar gran parte del contenido informativo de Occidente -en Oriente existen casos como el de China, quien bloquea el acceso a Facebook-. En realidad, la hipótesis que se plantea es que Facebook influye en los influencers. Los medios de comunicación y de opinión -aquí se analiza la prensa- dependen de sus canales de distribución, siendo Facebook uno de los más relevantes. Y si el canal de distribución modifica sus reglas -sus algoritmos- cada medio tendrá que adaptarse para conseguir visibilidad y mantener su influencia. Quien no se adapte al algoritmo, desaparecerá de Facebook, lo que es una manipulación mediante la invisibilización.

Para describir el poder de influencia de Facebook se ha optado por analizar sus entrañas: estudiar a fondo su "centro de mando comunicacional"; es decir, el "*News Feed* [el proveedor y actualizador de noticias para cada usuario concreto]". Y también cómo funcionan algoritmos de selección usando una metodología inspirada en la "ingeniería inversa": cartografiar los efectos de los algoritmos sobre una plataforma (en este caso Facebook) y, a partir de ahí, fueron desarrolladas predicciones de cómo pueden estar diseñados esos algoritmos y qué puede haber tras ese diseño.

Como bien recalca el estado de la cuestión el *News Feed* tiene un impacto gigantesco a nivel social -es una de las características que catapultó el éxito de Facebook-, económico -es responsable de gran parte de los ingresos publicitarios de Facebook- y tecnológico -muchísimas redes sociales se inspiraron en el *News Feed* para dar lugar a sus propios ecosistemas-. Por supuesto esto tiene un impacto notorio sobre la distribución del contenido

[30] https://learningenglish.voanews.com/a/online-privacy-concerns-raised-after-abortion-case/6697611.html

[31] https://www.msn.com/en-us/health/medical/colorado-mother-sues-facebook-alleges-daughter-s-addiction-to-platform-has-caused-mental-health-problems/ar-AA10pla2?li=BBnba9O

noticioso ya que por primera vez en la historia de Internet comenzamos a presenciar la centralización de la distribución de la información -muy lejos de la democratización que muchos anhelaron ver durante el desarrollo de Internet-; ahora Facebook se había convertido en una de las principales fuentes noticiosas y era precisamente Facebook quien a través de sus algoritmos decidía quién vería qué.

Facebook utiliza el *Machine Learning*, la disciplina científica que utiliza información y algoritmos matemáticos para crear sistemas que aprenden automáticamente, con el fin de puntuar las publicaciones y entregarlas a los usuarios a través de *News Feed* (Facebook, n.d.). Por tanto, la exposición -y jerarquización- de un contenido depende de los resultados de estos algoritmos. Estos filtros se basan en distintos puntos de información (de ahora en adelante "señales") -antigüedad de la publicación, interacciones, fuente, etc.- presentes en el contenido per se y en los objetos del *Social Graph* de Facebook.

Comprender qué es el Social Graph es fundamental para entender a Facebook y su *News Feed*. El *Social Graph* es un concepto tecnológico basado en la teoría de los grafos -rama de las matemáticas que combina álgebra, probabilidad y geometría-. El *Social Graph* es un mapa global de cómo cada usuario se relaciona dentro de Facebook: "¿siguen sus contenidos? ¿Tienen una relación? ¿Son amigos?, etc. (Fitzpatrick, 2007)". Facebook utiliza Social Graph para cumplir uno de sus principales intereses: rastrear todas las acciones que realizan los usuarios y disponer de un registro de las interacciones y conexiones sociales de cada usuario de Facebook y de éstos entre sí, (Dickinson, 2012). Es decir, *Social Graph* es una herramienta imprescindible para el correcto funcionamiento de *News Feed* al ser la fuente de información que Facebook construye constantemente y de la que los algoritmos de *News Feed* se nutren para calcular y hallar la información que más interesa a cada usuario.

En este sentido, Facebook usa un funcionamiento basado en retroalimentación circular: los usuarios interactúan con las publicaciones presentes en el *News Feed*, generando información que es albergada en el *Social Graph* y utilizada por los algoritmos para posicionar contenidos que aumenten la interacción de cada usuario. Es un círculo cerrado en donde el objetivo central es mantener la atención constante del usuario.

Facebook tiene el alcance, la atención y la información de 2.380.000.000 de usuarios, (Facebook, 2019). Ninguna otra red social la iguala en cuanto a volumen de audiencias e información. Es el mayor influencer de la historia. Con este escenario es importante preguntarse: ¿el funcionamiento de los algoritmos asociados a *News Feed* y las decisiones tomadas por Facebook para premiar o castigar determinados contenidos influyen directamente sobre el tipo de contenidos creados y/o distribuidos? Para responder a esta hipótesis se ha investigado los cambios aplicados por Facebook en su *News Feed* de manera que permita cartografiarlos y visualizarlos cronológicamente agrupando cada año en tres grandes grupos: "Nacimiento y transformación", "Manipulación mediante la invisibilización" y "*Fake news*, filtros y control de la información".

Tabla 1
Definiciones de indicadores en Facebook.

Indicador	Definición
Me gusta	Una interacción del usuario cuando éste hace clic en el botón "me gusta" de una publicación.
Compartir	Una interacción del usuario cuando éste hace clic en el botón "compartir" de una publicación.
Comentar	Una interacción del usuario cuando éste deja un comentario en una publicación.
Publicación	Un contenido publicado por un usuario en su propio perfil, página o grupo en formato de texto, video, enlace, imagen, galería, etc.
Interacciones por publicación	Es el número de interacciones totales (me gusta, compartir y comentar) adquiridos por un grupo de publicaciones dividido por el número total de publicaciones.
Historias	Publicaciones basadas en las interacciones de los amigos de un usuario
Fan Page	Perfil de página en Facebook que puede pertenecer a una institución, empresa, periódico, celebridad, etc. No puede tener amigos.

Indicador	Definición
Hoax	Un contenido deliberadamente falso que es presentado como si fuera verdad, puede ser considerado como una forma de *fake news*.
Viral	Contenido que se propaga fácilmente de forma orgánica entre usuarios.
Publicaciones Click-bait	Publicaciones diseñadas para llamar la atención de un usuario y obtener un clic en su contenido, todo mediante titulares llamativos que no son respaldados por un contenido de calidad.
engagement baiting	Es una técnica que promueve las interacciones en Facebook usando contenido sin valor y engañoso
Llamada a la acción	Frases o expresiones diseñadas para promover una acción por parte del usuario. Usualmente son frases como "clic aquí", "comprar ahora", "conócenos", etc.
Cloacking	Técnica utilizada para engañar a Facebook mostrándole un contenido distinto en una URL al que observan los usuarios.
Publicador	Caja donde cada usuario crea un post.
Exposición pagada	Invertir dinero para aparecer como anunciante en el *News Feed*
Exposición orgánica	Ser visible de forma gratuita y por selección de los algoritmos

Fuente: Elaboración propia.

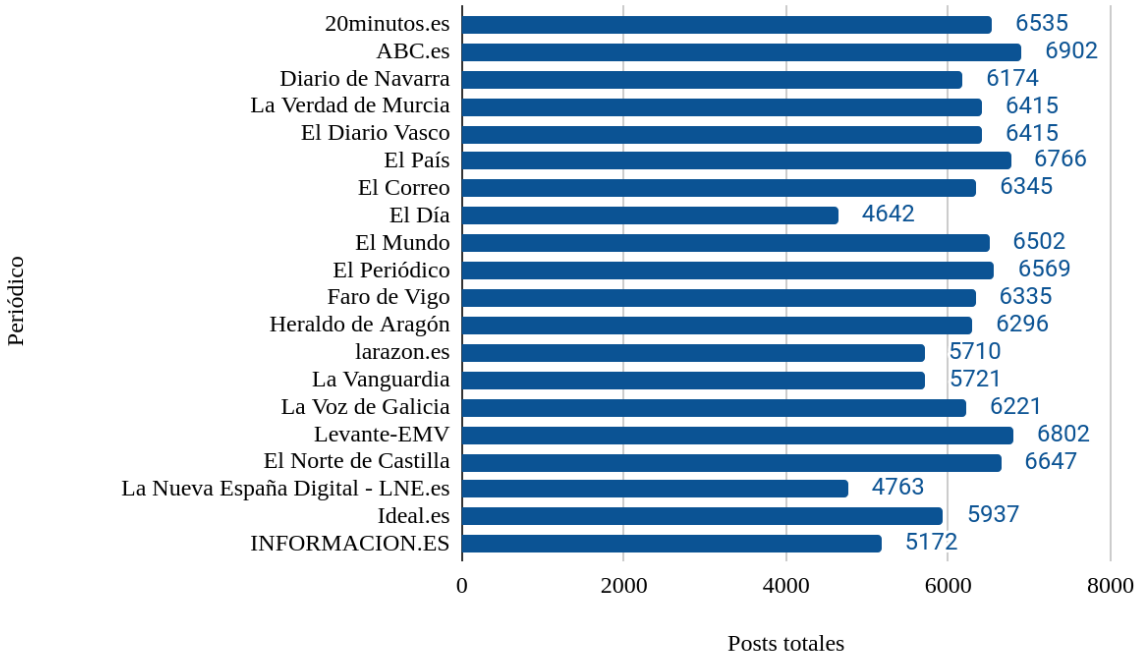
Metodología

Este artículo tiene dos partes diferenciadas. La primera describe cronológicamente los cambios del *News Feed* de Facebook incluyendo sus algoritmos. Este análisis solo incluye fuentes oficiales y verificables. Están muy desperdigadas y el objetivo es reunirlas. Fue creada una cartografía anual de los cambios que tienen una influencia explícita sobre el

News Feed, excluyendo cualquier otro que no tuviese esa relación. La segunda parte ha resultado un trabajo ingente pues han sido recolectadas todas las publicaciones disponibles mediante la Graph API de Facebook durante más de 10 años (de 2008 a 2019). Se seleccionaron los 20 periódicos españoles con mayor audiencia según el Estudio General de Medios (EGM) de febrero a noviembre de 2008. Fue elegido 2008 porque es la primera fecha de datos -febrero de 2008- que ofrece la *Graph API* de Facebook. Además de audiencia, los periódicos, obviamente, debían de tener perfil en Facebook y seguir activos en 2019. Los periódicos seleccionados y su respectivo número total de publicaciones incluidas en el análisis son: *El País* (6.766), *20 Minutos* (6.535), *El Mundo* (6.502), *La Vanguardia* (5.721), *El Periódico* (6.569), *ABC* (6.902), *La Voz de Galicia* (6.221), *El Correo* (6.345), *La Nueva España* (4.763), *Heraldo de Aragón* (6.296), *Levante El Mercantil Valenciano* (6.802), *Faro de Vigo* (6.335), *La Razón* (5.710), *La Verdad* (6.415), *El Diario Vasco* (6.415), *Diario Información* (5.172), *El Norte de Castilla* (6.647), *El Día* (4.642), *Ideal* (5.937) y *Diario de Navarra* (6.174). La cantidad de publicaciones recolectada para cada periódico depende de la actividad que dicho periódico haya tenido en Facebook. No descartamos ninguna publicación obtenida mediante el *Graph API* de Facebook.

Figura 1

Periódicos incluídos en el estudio y el total de publicaciones descargadas por cada uno.



Fuente: Elaboración propia.

En total fueron recolectados más de 120.000 publicaciones (en concreto 122.869 post de Facebook) en donde las más antiguas datan de febrero de 2008 y las más recientes hasta octubre de 2019; es decir, se ha cubierto más de 10 años. Debe matizarse que la recolección de estas publicaciones no supone el total de las realizadas por todos los periódicos analizados desde el año 2008, sino que estas 122.869 publicaciones, durante 10 años, resulta de una muestra estadística determinada por Facebook al seleccionar las publicaciones más destacadas -principalmente en interacciones y alcance-, además esta información no difiere de la que cualquier otra herramienta como Hootsuite o similares dispone. Como se observa en la figura 1 aunque no existe una recolección de datos completamente uniforme entre periódicos -principalmente porque cada uno de ellos tiene una frecuencia de publicación diferente- sí existe una tendencia generalizada hacia las 6.000 publicaciones en total por cada página de Facebook, esto es debido a que no se ha dado un trato preferente a ninguno de los periódicos sino que se han descargado y compilado únicamente los datos que Facebook facilita a través de su *Graph API.*

Para determinar esta influencia tenemos en cuenta que su *Graph API* -protocolo de comunicación mediante el cual Facebook permite acceder a la información del *Social Graph*- permite obtener aproximadamente 600 publicaciones en un año, que corresponden a las 600 publicaciones más destacadas -esto se menciona en la documentación oficial de las APIs de Facebook publicada en su web-. Por este mismo motivo, los gráficos e interpretación de los datos no se basan en números absolutos, sino relativos y las publicaciones analizadas no son una selección aleatoria, son las publicaciones que Facebook ha considerado como más relevantes -respecto a interacciones y alcance-, lo cual se sabe da mayor importancia a nuestros resultados, pues es analizada sólo la información que llegó a tener un poder de influencia notable en Facebook. Los periódicos fueron seleccionados como influenciadores debido a tres motivos principales:

Como periódicos, su uso habitual de las redes sociales es para la adquisición e influencia de audiencias, por lo que sacar el mayor provecho del *News Feed* es uno de sus mayores intereses, lo que significa que su comportamiento debería adaptarse, en teoría, a las actualizaciones del algoritmo de *News Feed*.

Facebook considera a las noticias como parte fundamental del contenido dentro del *News Feed* al ser un contenido de alta relevancia e influencia, por este motivo los periódicos pueden ser más sensibles a determinadas actualizaciones.

Al ser todos periódicos de España tiene sentido encontrar comportamientos comunes que reflejan la influencia de los algoritmos sobre sus comunicaciones en redes sociales.

Una vez recolectada la información, con los datos totales, fueron agrupados por trimestres y por distintas dimensiones de datos y métricas disponibles de cada publicación como: fecha de creación, hora de creación, fecha de actualización, hora de actualización, nombre de perfil, descripción, ID de post, tipo de post, enlace del post, tipo de contenido, historia, URL de imagen, mensaje, lugar, fuente del contenido, enlace en el post, aplicación utilizada para publicar, etiquetas de historia, me gusta, reacciones, comentarios y compartir.

Cronología de cambios destacados en la distribución de contenidos dentro de *News Feed* y sus puntos de poder e influencia

Los cambios históricos dentro de los algoritmos de *News Feed* tienen varias vertientes, algunas son tecnológicas y soportan la mejora de los algoritmos actuales o la implementación de nuevas técnicas de *Machine Learning* o *Inteligencia Artificial*. Y otras subjetivas que corresponden a la visión de Facebook como red social como puede ser el ajuste de los algoritmos para detectar material violento o dar prioridad a determinados formatos de contenido como el video. Cada cambio en los algoritmos del *News Feed* de Facebook supone toda una revolución en todos los medios del mundo.

Nacimiento y transformación de News Feed

Facebook lanza *News Feed* el 6 de septiembre de 2006 (D'Onfro 2016) convirtiéndose en la fuente de información de lo que ocurre en Facebook para un círculo de amigos. Para fortalecer esta herramienta en 2007 Facebook crea un algoritmo capaz de recolectar la información de la actividad de cada usuario y además lanza por primera vez el uso de esta red social en dispositivos móviles. Como grandes hitos vemos que en este año también crean los *Network Portals*, unos agregadores de información que facilitan a los usuarios el comprender y compartir información sobre sus entornos de trabajo, estudio, entretenimiento, etc. Facebook impulsa más de 47.000 agregadores en distintos conjuntos de países, centros educativos y empresas (Facebook, 2007). Esto implica que los usuarios comparten mayor información privada -o más bien, Facebook encuentra un método para empezar a trazar con mayor detalle a sus usuarios-.

En Facebook el usuario y su información son el producto, con lo cual al aumentar el volumen de información extraída de los usuarios decidieron dar lugar a los primeros servicios de anuncios publicitarios, los que a su vez influyen sobre la distribución de contenidos. Este servicio involucra alianzas con Microsoft y otras 12 compañías importantes del sector tecnológico (Facebook, 2007).

En 2007 surgen teorías y conceptos asociadas al funcionamiento de *News Feed* como el "*NFO*" (*News Feed* Optimization), una actividad hecha para manipular los algoritmos del *News Feed* para exponer contenidos (un post) a un grupo de personas específicas (una

audiencia). La industria publicitaria reconoce las señales utilizadas por Facebook para filtrar y exponer contenidos en el *News Feed* y la importancia de explotar esa información para obtener mayor visibilidad. Entre esos datos destacan: a) la capacidad de enganchar (publicaciones con un alto nivel de interacción); y b) la posibilidad de pagar por exponer un anuncio (o "*NFM*", *News Feed* Marketing) (Adweek, 2007).

Estas técnicas y estudios inician una carrera por la máxima visibilidad en el *News Feed*, todo debido al éxito de esta herramienta en el aumento de audiencias para Facebook. Todos querían hacer todo lo que los algoritmos pidieran con el fin de lograr visibilidad. La consecuencia de esto fue el aumento de las publicaciones disponibles en el *News Feed*, hecho que explica Justin Rosentein, ingeniero de software en Facebook, afirmando que sólo poco más del 0.2 % de las publicaciones consideradas por el algoritmo son expuestas. Todo es debido a unos "principios generales" con los que cada publicación es calificada, calificación que es adaptada y retroalimentada según el comportamiento de cada usuario (Adweek 2007).

Hacia 2008 Facebook renueva el diseño del *News Feed* y mejora sus funciones al introducir un nuevo publicador que permite crear publicaciones de fotografía, video o texto (Facebook, 2008). Esto implica que el formato -foto, texto, vídeo- va a va a ser importante para qué será visible o no dentro del *News Feed*.

Los avances técnicos del *News Feed* incrementan notoriamente en 2009 con la compra de FriendFeed -un agregador de información en la red que incluía noticias, redes sociales, blogs, etc.-. Esta compra influye en la estructura de los productos de Facebook (Grove 2009), en especial del *News Feed*, ya que replican características que ya existen en FriendFeed: comentarios en publicaciones, dar "me gusta" y reestructurar *News Feed* hacia una denominada "real-time homepage", es decir, un *News Feed* que a partir de este momento muestra información en tiempo real y no en intervalos de 10 o 15 minutos como era hasta este momento (Ostrow 2009).

Estos cambios tecnológicos también influyeron sobre el tipo de información que los usuarios consumían, *News Feed* pasó por una transformación en la que dejó de ser una herramienta dedicada a compartir publicaciones de amigos y familia a incluir contenidos informativos (en especial gracias a las nuevas características de "like" y comentarios)

(Ostrow 2009). Esto fortaleció la presencia de un contenido comercial en *News Feed* y aceleró la carrera por la visibilidad, ahora marcas, periódicos e instituciones competían con mayor frecuencia para lograr hacerse con un espacio en el *News Feed* de la mayor cantidad de usuarios posibles. Todo ello también obligó a Facebook a mejorar sus filtros de contenido al incluir en su análisis algorítmico factores de relevancia, afinidad y antigüedad. Desde octubre de 2009 las publicaciones con mayor nivel de interacción tienen mayor visibilización (Bradley 2009). Este hecho es muy relevante y estará asociado años después a la aparición de las *fake news* y distribución de contenidos falsos.

Para 2010 ocurre uno de los cambios más importantes para Facebook, el lanzamiento de *Graph API* como método de interacción con el *Social Graph* para inyectar o extraer información de usuarios o páginas en Facebook. También aparecen las primeras versiones de los plugins sociales (permite conectar cualquier web con Facebook). Esta estrategia es pública en la conferencia F8 (la conferencia anual de Facebook para programadores informáticos) (McCarthy, 2010). Esto es todo el fundamento tecnológico que Facebook necesita para obtener información de todos los usuarios y aplicar técnicas de *Big Data* -grandes volúmenes de datos- para explotar información privada con fines comerciales.

Un hecho teórico importante lo lidera Widman (n.d.) al desarrollar el concepto de *EdgeRank* para explicar cómo Facebook elige los contenidos que aparecen en el *News Feed*. *EdgeRank* es una explicación simple y precisa de cómo funciona el algoritmo de *News Feed*. Esta teoría se hará obsoleta con el tiempo, pero la reconocemos como una de las primeras explicaciones más coherentes e importantes sobre *News Feed*.

Todos los cambios de *News Feed*, desde su lanzamiento en 2006 hasta el uso de *Graph API* en 2010 componen todos los fundamentos técnicos de esta herramienta. Por ello consideramos esta etapa como el "nacimiento" de *News Feed* ya que no solo se trata de publicar esta herramienta, también es necesario incluir todas las tecnologías que fundamentan su uso hoy en día. En 2006 *News Feed* solo era un concepto, una idea nueva con un enorme potencial, y en 2010 terminó siendo una potente herramienta publicitaria y de *Big Data*, con un nivel notorio de influencia sobre empresas e instituciones en donde todos obedecieron las reglas de sus algoritmos -creando incluso teorías sobre cómo explotar

los algoritmos y sacar provecho de ellos- con una capacidad sin precedentes de recolectar información de usuarios a nivel de interacción.

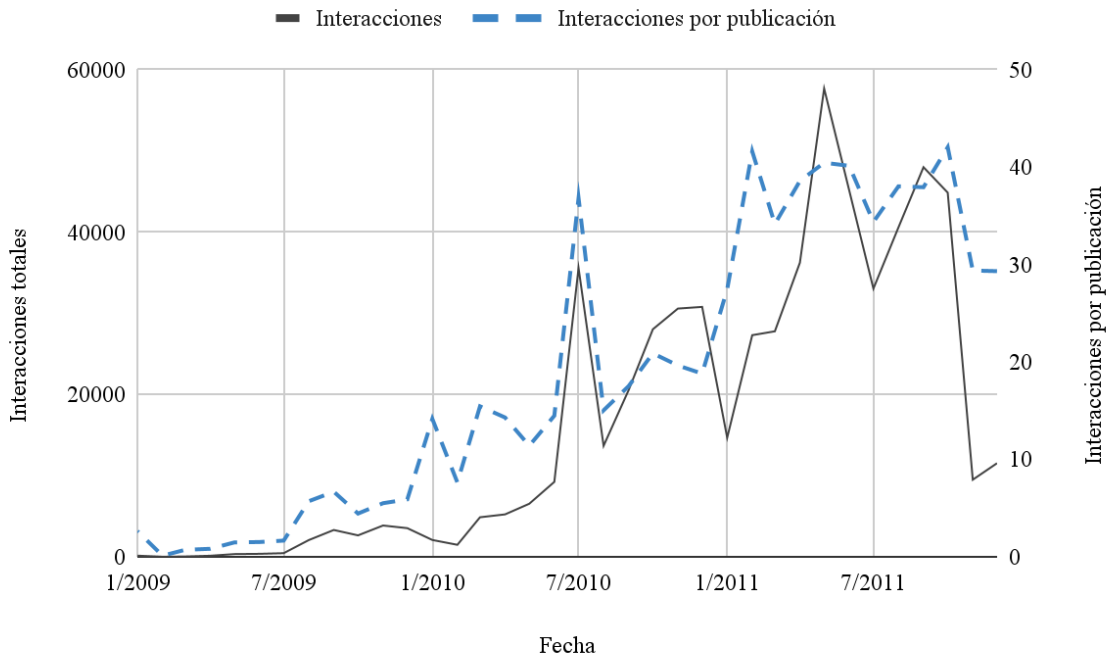
La etapa de "manipulación mediante invisibilización"

Para este momento, Facebook conoce el poder de influencia de *News Feed*, por lo que ejecuta estrategias de "manipulación mediante la invisibilización", métodos sutiles y efectivos con los cuales obliga a empresas, anunciantes, periódicos y cualquier otro propietario de una Fan Page a seguir las reglas de sus algoritmos o de lo contrario serán invisibles dentro del *News Feed*. Mientras en años anteriores todo estaba concentrado en la construcción de unos fundamentos técnicos sólidos, en 2011 se empiezan a notar las capacidades de *News Feed* para manipular tendencias en comunicación.

Facebook otorga en 2011 mayor visibilidad al contenido noticioso y no únicamente a publicaciones de amigos, familia y perfiles a los cuales un usuario se encuentra suscrito (Tonkelowitz, 2011). Facebook se transforma de simple red social a medio de comunicación en toda regla al jerarquizar las noticias más importantes. Estos cambios son complementados por un nuevo botón de suscripción que permite a cada usuario "suscribirse" a otro para ver sus publicaciones en el *News Feed* (Parker 2011), a su vez cada usuario tiene la posibilidad de seleccionar dos tipos de suscripciones: casi todas las publicaciones o solo las publicaciones importantes (Rait 2011). A pesar de que Facebook cede cierto nivel de control a los usuarios, son los algoritmos el principal filtro de contenidos. Con el paso del tiempo incluso las suscripciones de un usuario podrían no estar en el *News Feed* si los algoritmos lo deciden.

Figura 2

Interacciones totales e interacciones por cada publicación de los 20 periódicos mencionados desde enero de 2009 hasta diciembre de 2011



Fuente: Elaboración propia.

Al analizar los resultados en interacción de los periódicos españoles seleccionados desde 2009 hasta 2011 observamos un crecimiento de las interacciones por publicación durante 2011 -año en el que *News Feed* incrementa la visibilidad de las noticias- lo que significa que los periódicos necesitaron menor esfuerzo para incrementar sus niveles de influencia durante este periodo. Esto demuestra cómo los cambios del algoritmo del *News Feed* influyen de forma clara sobre los resultados de las interacciones y que en esta época supuso un incremento en participación desde 2009, realidad que también reflejan en algunos estudios como mencionan Lysak *et al* (2012):

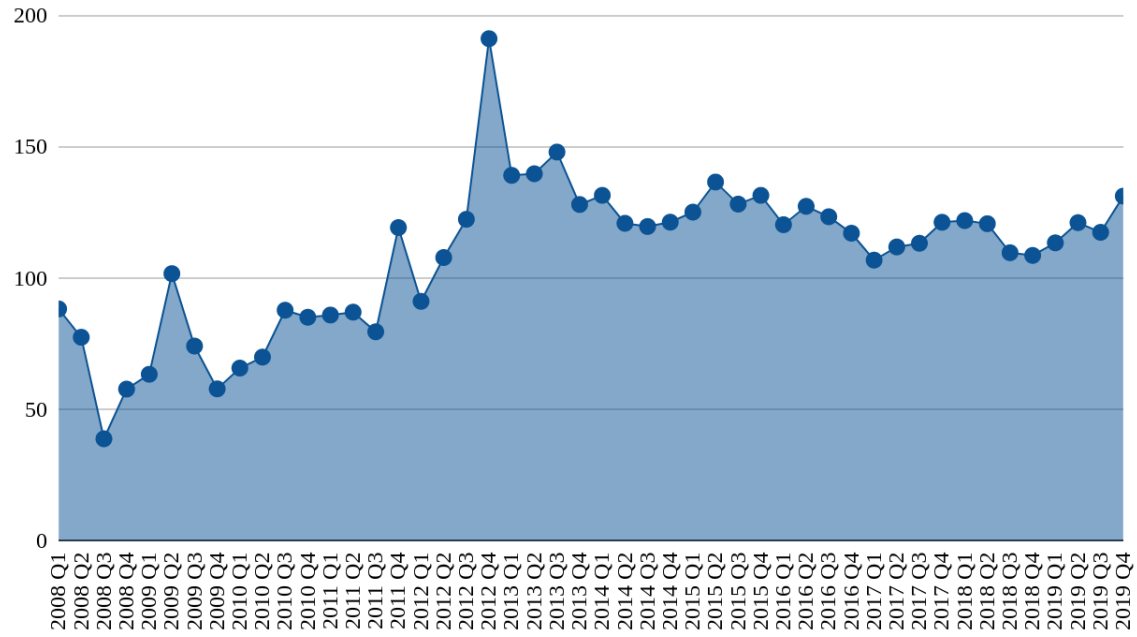
> "...los reporteros y la gerencia de la sala de redacción están utilizando las redes sociales para aumentar su perfil en las comunidades que cubren [...] indicando que regularmente publican enlaces a sus noticias en Facebook, y casi la misma cantidad en Twitter para el mismo propósito, las redes sociales ya se han convertido en una valiosa herramienta de promoción. [...] creen firmemente que el uso de las redes

sociales ha mejorado tanto el reportaje de historias locales como la interacción de sus salas de redacción con los televidentes. Esto podría conducir lógicamente a un estudio de cómo las redes sociales influyen en las decisiones editoriales...”

(p. 203)

Figura 3

Longitud promedio en caracteres de publicaciones en Facebook.



Fuente: Elaboración propia.

Como evidencia la Figura 3 los cambios de Facebook también alteran la propia longitud del mensaje publicado. En este caso los periódicos iniciaron hacia el año 2008 con mensajes cortos inferiores a los 100 caracteres -y en ocasiones inferiores a los 50- llegando casi hasta los 200 en 2013. En efecto esta tendencia se vio cortada y ahora la longitud media se ubica entre los 100 y los 150 caracteres.

El poder de influencia de Facebook se debe a su enorme audiencia y por ello *News Feed* existe para entretener a las masas. Ello implica que en 2012 Facebook utiliza técnicas de *Big Data* para extraer información de las interacciones entre usuarios para entender cómo distribuyen y aceptan las noticias que les gustan o disgustan, esto lo vemos en un estudio de Bakshy (2012): “Para comprender cómo las redes sociales afectan la difusión de la

información, utilizamos variaciones aleatorias en el *News Feed* para determinar la probabilidad de que una persona comparta contenido web si vio o no el contenido compartido por sus amigos." (parr. 13). Esto demuestra el interés de Facebook por el tipo de contenido, la ideología política, el contenido noticioso y su relación con las conexiones sociales entre usuarios. Es un reconocimiento implícito de Facebook de que su supremacía se debe en parte gracias a su capacidad de entender a su gran audiencia.

También surgen nuevas características importantes como la posibilidad de convertir "amigos" en "conocidos" para ver sus contenidos con menor frecuencia (Coens, 2012). Esto propicia la aparición del fenómeno de las "cámaras de resonancia" (Elías, 2019) que favorece que los usuarios solo interactúen con aquellos que piensan de forma similar y que también influye en la distribución de las *fake news*. Aspecto que curiosamente contradice la sugerencia de Bakshy (2012) de que Facebook no es una "cámara de resonancia".

A partir de 2012 los anunciantes pueden publicar anuncios dentro del *News Feed* (Protalinski, 2012). Esto trae conceptos como la "exposición pagada" -pagar para que ciertas publicaciones tengan mayor visibilidad en Facebook- y la "exposición orgánica" -visibilidad que recibe una publicación por decisión de los algoritmos de Facebook-. Tras la aparición de esta nueva modalidad de anunciar surgen distintos análisis que demuestran una notable disminución del alcance de las publicaciones asociadas a empresas en Facebook (Peterson, 2012). Este es un precedente importante que demuestra la evidente manipulación de Facebook sobre los resultados del *News Feed* para aumentar sus beneficios económicos ya que la única forma de obtener la misma exposición que se obtenía antes de forma "gratuita" era ahora pagando anuncios.

Los cambios de 2013 involucran la cronología de las publicaciones con una actualización llamada "*Story Bumping*" (o saltos de historias) que consiste en exponer publicaciones del *News Feed* en orden según su importancia para el usuario y no necesariamente en un orden cronológico (Backstrom, 2013). Otro cambio llamado "*Last Actor*" que otorga mayor relevancia a las páginas o personas con las que un usuario interactuó en sus últimas 50 interacciones (Lafferty, 2013). Estas modificaciones influirán mucho sobre los contenidos al reforzar la existencia de una "*cámara de resonancia*" y al aumentar aún más la exposición de publicaciones (en muchos casos noticias) que tengan una alta popularidad,

además de facilitar la "manipulación mediante la invisibilización" ya que conceptos como la relevancia son subjetivos y pueden cambiar en el tiempo: tienes que ser relevante para Facebook para no ser invisible. Este año Facebook lanza los "*hashtags*", Warman, M. (2013). Una función que permite agrupar temas etiquetados mediante el signo "#". Esto posibilita agrupar y encontrar en el buscador de Facebook de una forma más simple los temas tratados en distintas publicaciones. Es sin duda una característica de uso obligatorio para los medios de comunicación.

Durante los últimos meses de 2013 los algoritmos del *News Feed* cambian para exponer a los usuarios anuncios más relevantes (Wasserman, 2013). Ello implica que los factores de posicionamiento en el *News Feed* tienen más fuerza no solo para las publicaciones orgánicas -de exposición gratuita-, sino también para las publicaciones pagadas -anuncios-. Estos cambios, que parecerían inocentes, tendrán en el futuro importantes repercusiones como el escándalo de Cambridge Analytica (explotación de datos de usuarios por los políticos), ya que es la misma información del usuario la que permite a Facebook establecer la relevancia de los anuncios. El usuario es el producto que Facebook vende a sus anunciantes y el arma con el que manipula a los medios. A su vez no existe limitación de contenido político o de naturaleza propagandística en estos anuncios.

La manipulación de Facebook utilizando la invisibilización se profundiza a partir de 2014 al aumentar el número de contenidos que cada usuario consume al igual que el número de interacciones que hacen mediante distintas acciones concretas: el lanzamiento de "trending topics" sobre todo tipo de temas y contenidos, incluyendo las noticias (Struhar, 2014), considerar como más importantes las publicaciones que rápidamente obtienen muchas interacciones, (Owens y Vickrey, 2014) y publicar automáticamente "historias" ocasionando que todos los usuarios vean en su *News Feed* publicaciones de personas o páginas para las cuales no manifestaron ningún interés previo (Song, 2014). Estos cambios se concentran en premiar la "viralidad" -la distribución de contenido de un usuario a otro- y la interacción acelerada de los usuarios. Las publicaciones que no alcanzan ciertos niveles de interacción, aunque sean un buen contenido, serán invisibilizadas frente a las publicaciones que sus algoritmos premia. Estos cambios dan a entender que Facebook

conoce el potencial de *News Feed* como fuente de información y actualidad para los usuarios y están diseñados para extraer beneficios de este potencial.

Esta idea de "*News Feed*" como herramienta de información y divulgación noticiosa se refuerza con *FB Newswire*; un servicio que permite a periódicos y periodistas incluir en sus medios material como fotografías o videos que se encuentran dentro de Facebook (Mitchell, 2014). También aparece *Facebook Media*, un servicio pensado para mejorar la experiencia de periodistas, marcas y famosos para que optimicen el impacto que generan en esta red social (Facebook, 2014). Estos cambios no son aleatorios, Facebook sabe perfectamente que los medios utilizan publicaciones en el *News Feed* para ganar audiencias, y al mismo tiempo saben que Facebook es una fuente de información para generar noticias. Facebook quiere tener a todos alineados bajo sus reglas de juego: si todos utilizan sus herramientas y se adaptan a las reglas de sus algoritmos su poder de influencia se verá fortalecido.

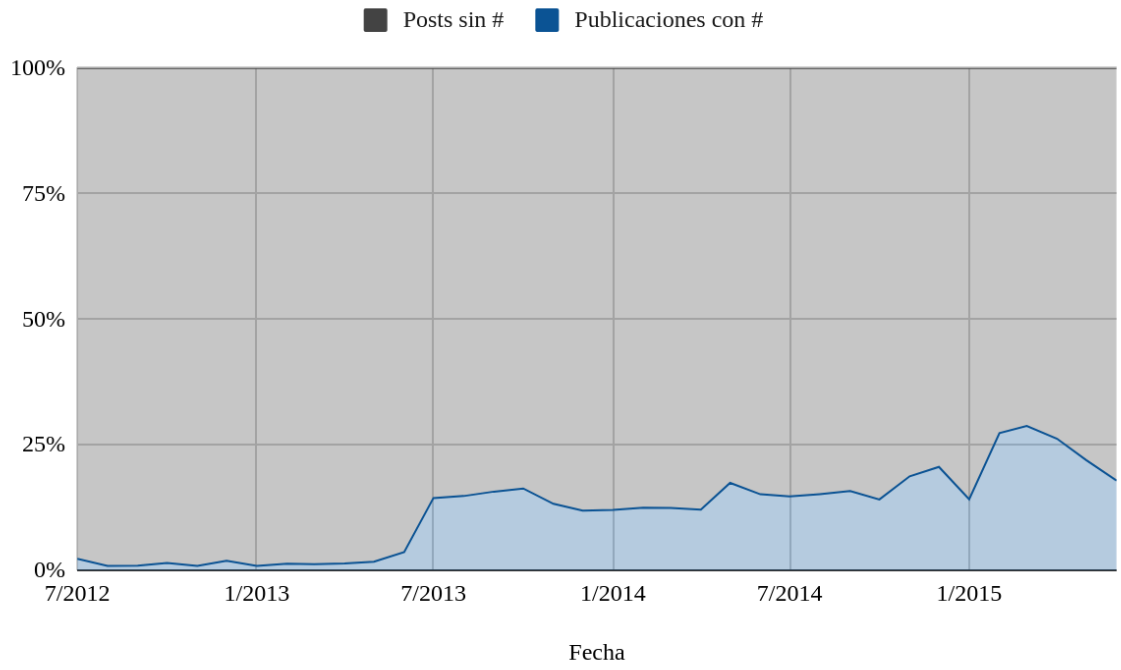
Los cambios anteriores son exitosos y aumentan la interacción de los usuarios, pero traen consecuencias con las que Facebook tiene que lidiar, como aplicar medidas correctivas al notar un aumento indeseado de memes y contenidos "*click-baiting*". No es ningún secreto que los algoritmos pueden ser engañados y en este caso los creadores de contenidos lograron tener un éxito temporal. Las medidas correctivas incluyen la medición de la permanencia de los usuarios en las páginas que visitan: si el tiempo de visita es alto entonces el contenido será considerado de alta calidad, de lo contrario sería castigado y reducida su exposición al ser posible que se trate de un contenido de "*click-baiting*" (El-Arini y Tang, 2014). También desde finales de 2013 Facebook quiere determinar qué contenidos tienen mayor interés para los usuarios, y decide que los memes tendrán menor relevancia en el *News Feed* (Kacholia y Ji, 2013)

Finalmente 2014 tiene cambios importantes en el funcionamiento de sus videos nativos con una serie de cambios concretos: Reproducirlos de forma automática y mejorar cómo son compartidos y visualizados (Simo, 2014) además de aplicar nuevos esfuerzos en medir señales importantes para conocer la relevancia de cada video, por ejemplo: el tiempo de permanencia de un usuario visualizando un video (Welch y Zhang, 2014). Estos cambios aplican únicamente para los videos nativos de Facebook y no para otras plataformas

externas como YouTube. Esto tiene repercusiones importantes sobre la distribución de contenidos ya que promueve la reproducción de videos en Facebook por encima de otros servicios, lo que evidencia un favoritismo por exponer videos de Facebook en el *News Feed* (O'Reilly, 2014). Estos cambios coinciden con la visión de Mark Zuckerberg de que en cinco años Facebook tendrá una mayor exposición de video (Miners, 2014).

Figura 4

Participación porcentual de las publicaciones con Hashtag en el total de publicaciones mensuales.



Fuente: Elaboración propia.

Los cambios del *News Feed* vistos entre 2012 y 2015 no son inocuos, la figura 4 expone explícitamente la influencia de Facebook sobre cómo lo medios producen contenidos. Es importante considerar dos aspectos de la gráfica, el primero es el cambio abrupto en la cantidad de publicaciones con *Hashtag*, pasando de participaciones inferiores al 2 % a más del 10 % en julio de 2013. Así mismo la existencia de los "*trending topics*" durante 2014 hizo que el número de publicaciones con *Hashtag* aumentara su participación porcentual mes a mes -ya que presuntamente el uso de un *Hashtag* puede aumentar la probabilidad de ganar exposición gracias a esta nueva característica de Facebook-, tendencia que se sostiene

hasta 2015. Sin lugar a dudas, los cambios de Facebook influyen en las técnicas que utilizan los periódicos para comunicarse.

Fake news, filtros y control de la información

Como ha sido observado, desde la creación del *News Feed* hasta el año 2014 todos sus cambios tienen como objetivo principal el aumento de las interacciones y la retención de los usuarios. Todos estos cambios hasta la fecha dieron a Facebook una audiencia de talla global, y con ello ciertas responsabilidades. A partir de 2015 se empieza a ver una serie de actualizaciones hechas para moderar la información noticiosa y a evitar la proliferación de *fake news*. La primera de ellas detecta "*hoaxes*" -un tipo de información falsa que puede definirse como un "rumor"- y reduce su presencia en el *News Feed*. Para lograr esto permite a los usuarios informar sobre aquellas publicaciones consideradas como dañinas de forma que si Facebook detecta un alto número de informes señalando a una publicación de ser un *hoax* la castigaría con una exposición reducida en el *News Feed* (Owens y Weinsberg, 2015). Este mecanismo no impedirá la proliferación de las noticias falsas en el futuro ya que asume que los usuarios conocen qué es y qué no es verdad, un aspecto difícil de diferenciar en especial cuando una *Fake New* está diseñada para engañar y manipular. Este cambio lo reconocemos como uno de los primeros mecanismos disponibles para eliminar la información dañina en el *News Feed*.

Reconocer y categorizar un *hoax* es difícil ya que existen otros contenidos diseñados para ser virales que aunque puedan parecer contenido malicioso no lo son necesariamente. Sabemos que en 2015 Facebook utiliza encuestas para comprender mejor este tipo de contenidos y saber si el contenido viral es algo que las personas desean consumir o si se trata de información falsa (Tas y Chiraphadhanakul, 2015). Este tipo de acciones demuestran una dificultad real de reconocer la información falsa del entretenimiento y más adelante estos esfuerzos no permitirán controlar adecuadamente la proliferación de noticias falsas.

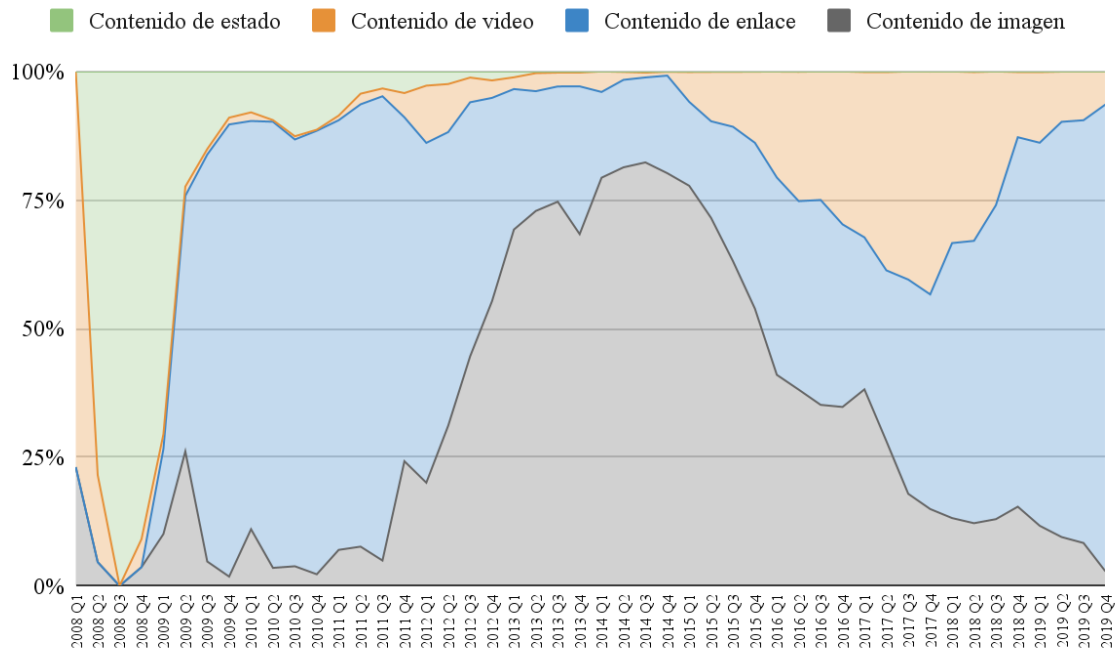
En 2015 Facebook aplica actualizaciones dedicadas a extraer en mayor detalle información sobre la interacción de los usuarios. En primer lugar Facebook empieza a valorar la permanencia en cada publicación al descubrir que los usuarios pese a tener un gran interés

en ciertas publicaciones no necesariamente interactúan con ellas; aunque sí permanecen un tiempo considerable viéndolas en pantalla. Por ello el tiempo de permanencia se convierte en una señal importante para determinar la relevancia de un contenido (Yu y Tas, 2015). Para las publicaciones de video consideran otras señales de relevancia: la activación de audio y la vista en pantalla completa (Wang y Zhuo, 2015). Los cambios de Facebook en el *News Feed* de priorizar los contenidos de video influirán sobre los periódicos: éstos aumentarán sus contenidos de vídeo para tener más visibilidad en el *News Feed*. Estos cambios junto a la preferencia de Facebook para que las publicaciones vistas en el *News Feed* sean las compartidas por familia y amigos del usuario frente a otras de diversas procedencias (Chowdhry, 2015) comienzan a presionar a los anunciantes para que inviertan dinero en publicidad para tener exposición. Cada vez se hace más importante gastar en Facebook para no ser invisible.

Como último cambio observado en 2015 tenemos las "reacciones", una nueva característica para que los usuarios expresen distintas interacciones: "me gusta", "me encanta", "me divierte", "me enfada", "me asombra" y "me entristece". Facebook utiliza estos datos para conocer al usuario (Tosswill, 2015). Estos datos pueden ser relevantes para técnicas de análisis de sentimiento y su uso posterior en hipersegmentación de audiencias. Usadas en partidos políticos, por ejemplo. (Elías, 2015).

Figura 5

Participación porcentual de las publicaciones según su tipo: estado, video, enlace o imagen.

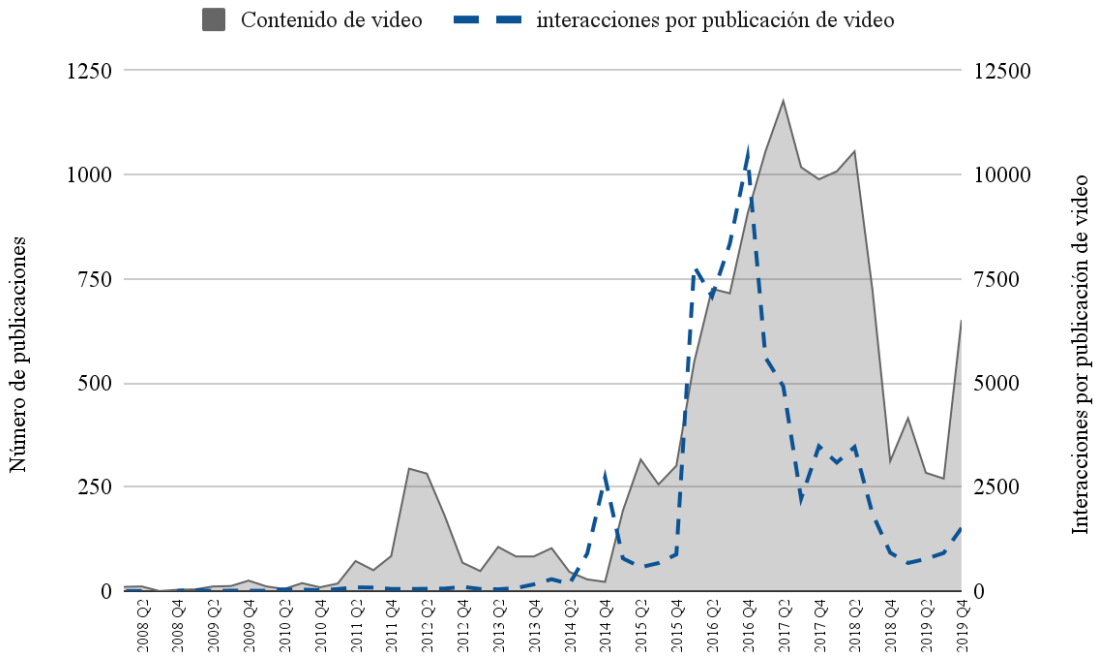


Fuente: Elaboración propia.

Desde el momento en que los algoritmos de *News Feed* favorecen la visibilidad de contenidos de video los periódicos publican con mayor frecuencia publicaciones de este tipo. Sin lugar a dudas esto es fruto de la *manipulación mediante la invisibilización*, después de todo si un periódico se negara a seguir las preferencias de los algoritmos sería invisible. Una audiencia tan grande como lo es Facebook no es algo que un periódico esté dispuesto a sacrificar solo por negarse a seguir estas reglas. Esta realidad es más evidente en la figura 5 con la cantidad total de interacciones en los contenidos de video.

Figura 6

Número de publicaciones de video vs interacciones por cada publicación de video.



Fuente: Elaboración propia.

Es importante observar en la figura 6 el retraso que existe entre la línea azul -interacciones por publicación de video- y la línea de área gris -contenido en formato de vídeo- ya que a medida que las interacciones por publicación de video aumentan, los periódicos reaccionan semanas después incrementando el número total de publicaciones con este formato. Así mismo, con la reducción de las interacciones por publicación de video el número de videos también es reducido. Esto confirma otra vez cómo los cambios algorítmicos del *News Feed* al favorecer ciertos formatos y despreciar otros influyen sobre los contenidos publicados por los periódicos. Nadie quiere ser invisible y la manipulación de Facebook se muestra efectiva.

El año 2016 sigue los mismos pasos de lo visto en 2015; Facebook continúa su persecución del contenido "*click-bait*" detectando frases y llamadas a la acción usadas de forma común por este tipo de contenidos, así como el reconocimiento de los dominios y sitios web que tienden a publicarlos. Facebook quiere anticiparse y detectar el contenido "*click-bait*" antes de que sea reportado por usuarios (Peysakhovich y Hendrix, 2016). Esta forma de actuar

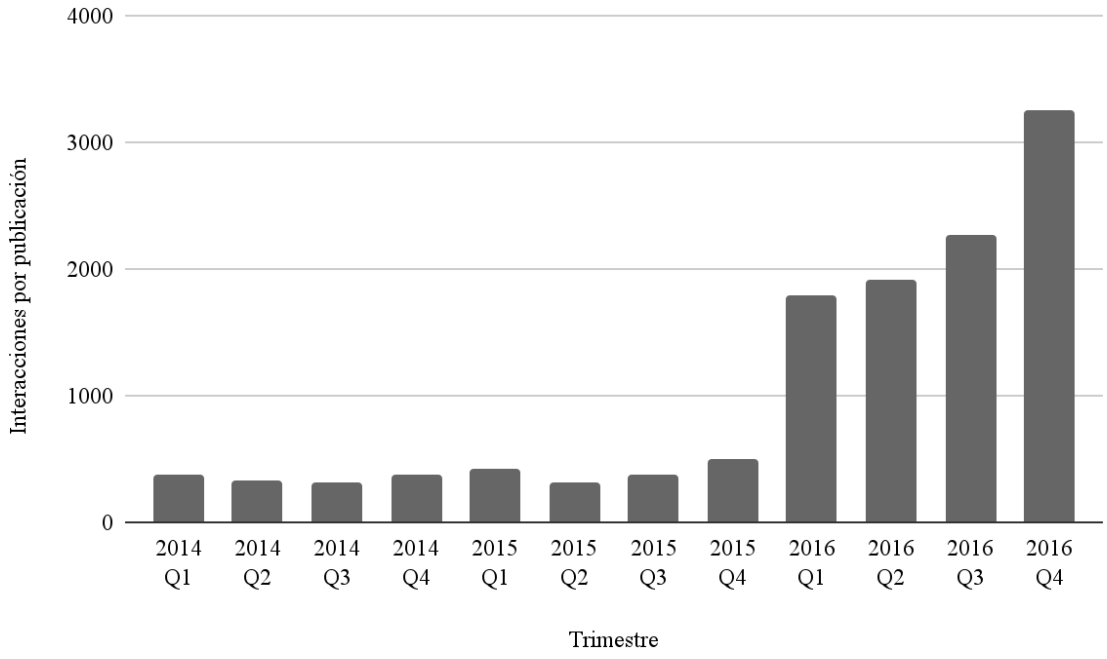
hace parte del tipo de acciones que Facebook podría tender a efectuar para combatir las *fake news*, aunque sigue sin ser una técnica efectiva para detectar el contenido desinformativo, únicamente es eficaz para encontrar el contenido de baja calidad.

En efecto, 2016 acentuó más los cambios vistos en 2015; por un lado, para *News Feed* son aún más importantes las publicaciones si son publicadas por amigos o familiares cercanos del usuario (Backstrom, 2016) y por otro lanza "*Feed Quality Panel*", una opción para que los usuarios otorguen un feedback sobre los contenidos visibles en su *News Feed* (Zhang y Chen, 2016). Estos cambios no hacen mucho eco sobre el comportamiento de Facebook en general y siguen la línea general trazada desde hace varios años: Facebook quiere que sus usuarios se entretengan en Facebook, y estos cambios están diseñados para lograrlo.

También observamos fenómenos benignos para los periódicos, aunque Facebook sabe que la invisibilización es un método efectivo de manipulación y también que el contenido noticioso beneficia la interacción y la permanencia de los usuarios. Por ello los algoritmos en 2016 priorizan publicaciones informativas y a su vez Facebook busca entender qué le importa al usuario desde el punto de vista informativo como entretenimiento (Mosseri, 2016). Esto trae una mayor exposición a los periódicos, pero también es el tipo de cambios que promueve la proliferación de *fake news* -dado que no es posible en este momento definir la verdad y tampoco existen mecanismos eficientes en el *News Feed* para reconocer fuentes fiables de información-. La figura 7 muestra cómo 2016 gracias a los cambios mencionados multiplicó la exposición e interacción que todos los periódicos recibieron.

Figura 7

Número de interacciones por cada publicación promedio en cada trimestre.

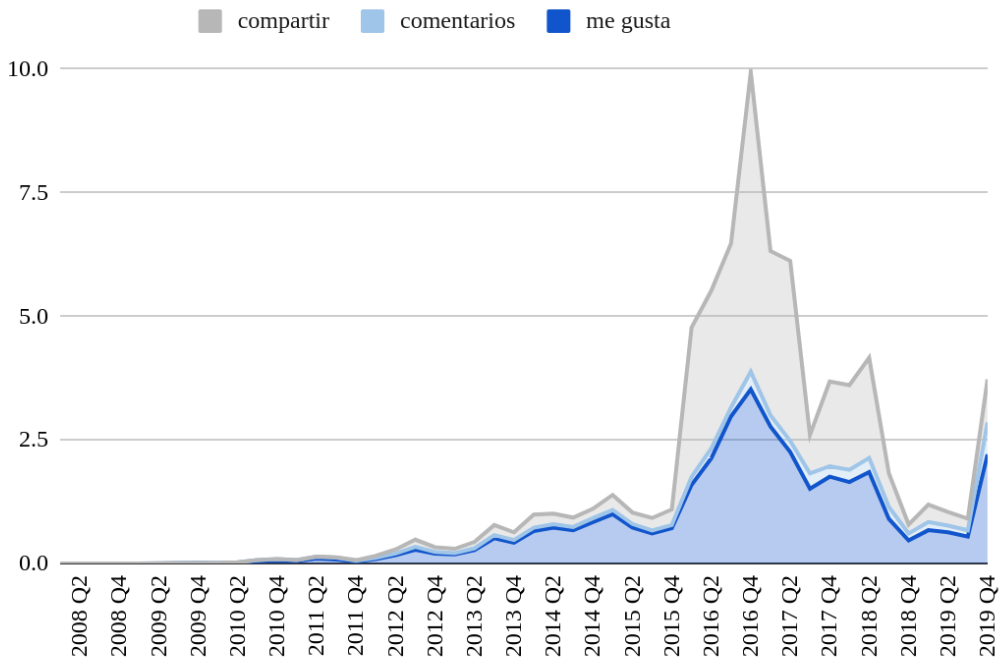


Fuente: Elaboración propia.

Esta gráfica da a entender que Facebook tiene completo control de lo que cada usuario ve. Si se asume que todos los periódicos de España optimizan sus publicaciones para enganchar a sus usuarios no se explica por qué desde 2014 la tendencia de interacciones por publicación es casi plana. No es sino hasta que Facebook decide que las noticias son importantes para los usuarios que las interacciones de los periódicos se multiplican y toman una tendencia claramente alcista. Aunque estudiosos del tema mencionaron la existencia de técnicas de *NFO* desde 2007, queda en evidencia que estas técnicas no son nada comparado con las decisiones unilaterales de Facebook sobre sus algoritmos.

Figura 8

Interacciones por trimestre dividido por tipo de interacción



Fuente: Elaboración propia.

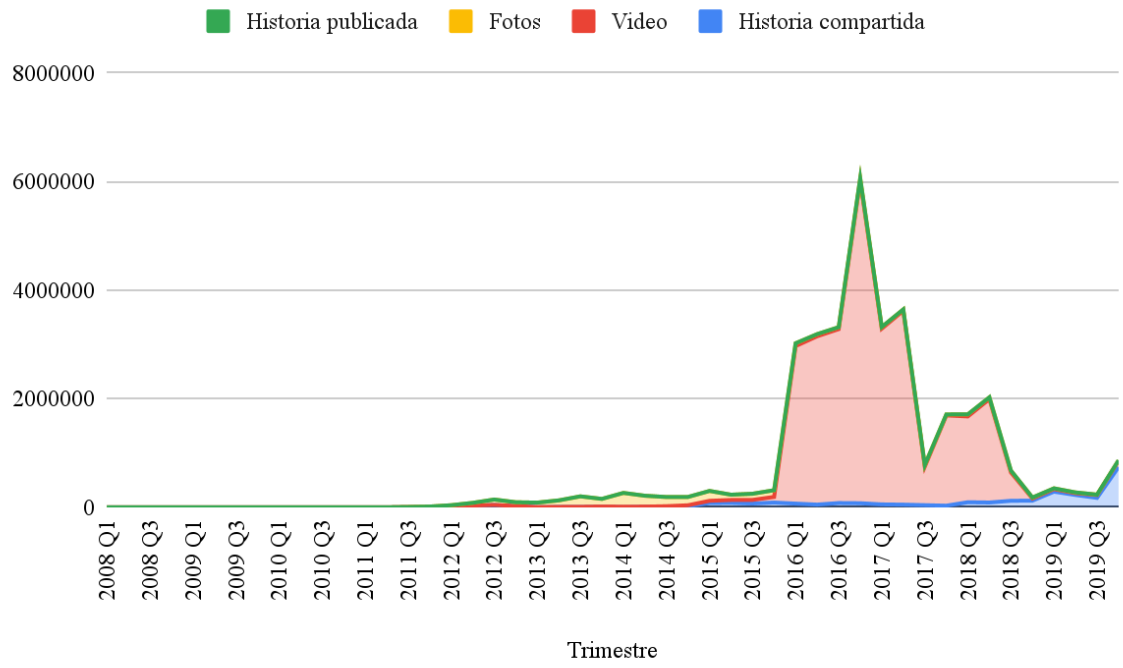
Uno de los aspectos llamativos de estos datos es el fuerte incremento en las interacciones de "compartir" en las publicaciones de los periódicos; es decir, es observable que las actualizaciones del año 2016 facilitaron y promovieron la exposición de los periódicos, no obstante estos datos evidencian lo que podría ser un "círculo virtuoso" de interacción y exposición que los periódicos obtuvieron temporalmente entre el año 2016 y 2018, ya que la interacción de compartir, a diferencia del "me gusta" y el "comentario", es un apalancador mucho más significativo del alcance en Facebook; es decir, cuando una publicación es compartida esta es vista por muchas más personas. Esto implica que las publicaciones de los periódicos en este periodo no solo obtuvieron más visibilidad por el

favoritismo que Facebook les otorgó debido a sus actualizaciones, también se debe a que los usuarios eran mucho más propensos a compartir estas publicaciones.

En este caso se ahonda sobre la causa de dicho incremento en el total de publicaciones compartidas y observamos en la Figura 9 que el gran incremento ocurre específicamente en las publicaciones de video.

Figura 9

"Compartir" por tipo de publicación y trimestre



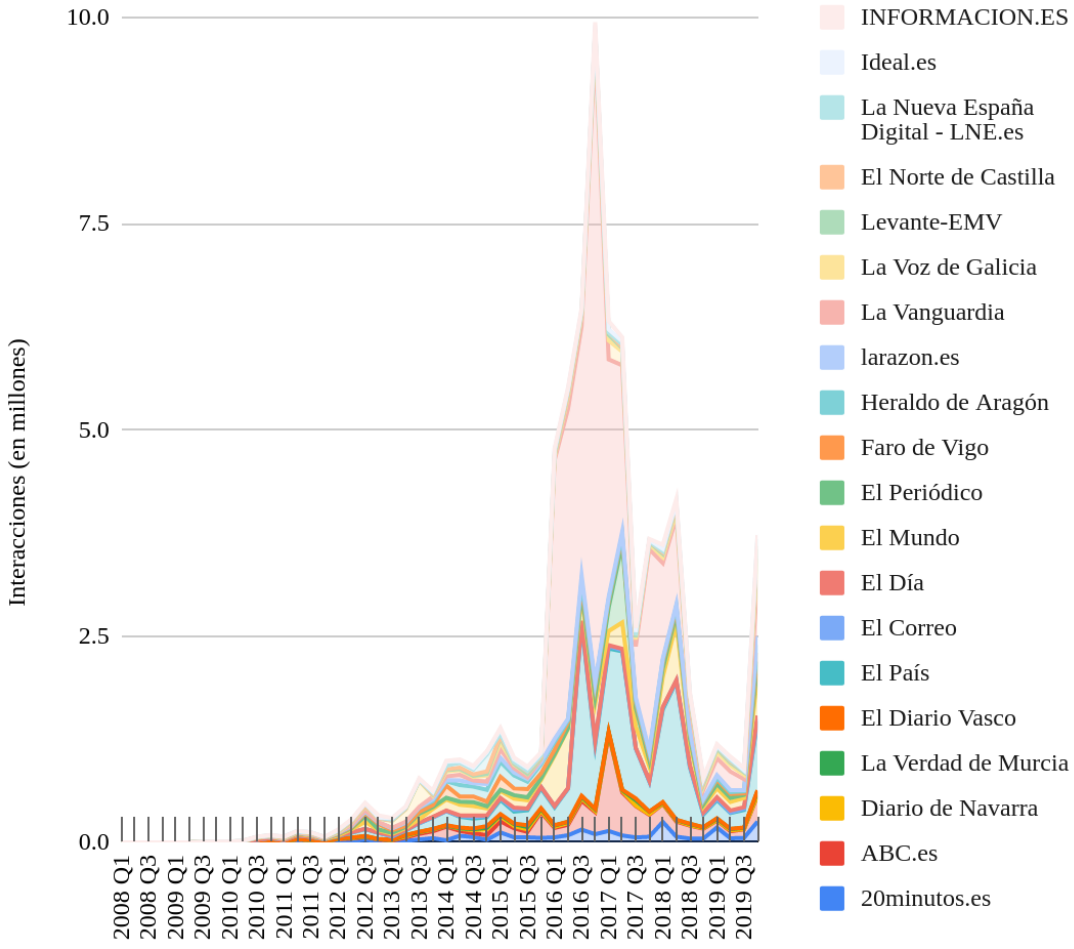
Fuente: Elaboración propia.

Aunque es difícil dilucidar la causa exacta por la que existe un repunte en las interacciones de "compartir" a partir del año 2016, la figura 9 sugiere que dicha causa está asociada principalmente a la prioridad que Facebook dio a los contenidos de video ya que son estos los que con una amplia diferencia recibieron este tipo de interacciones.

Otro aspecto a validar es la igualdad de comportamiento entre periódicos a que es posible que el comportamiento de un único periódico sea el responsable de la visible distorsión de datos.

Figura 10

Total de interacciones por periódico en cada trimestre.



Fuente: Elaboración propia.

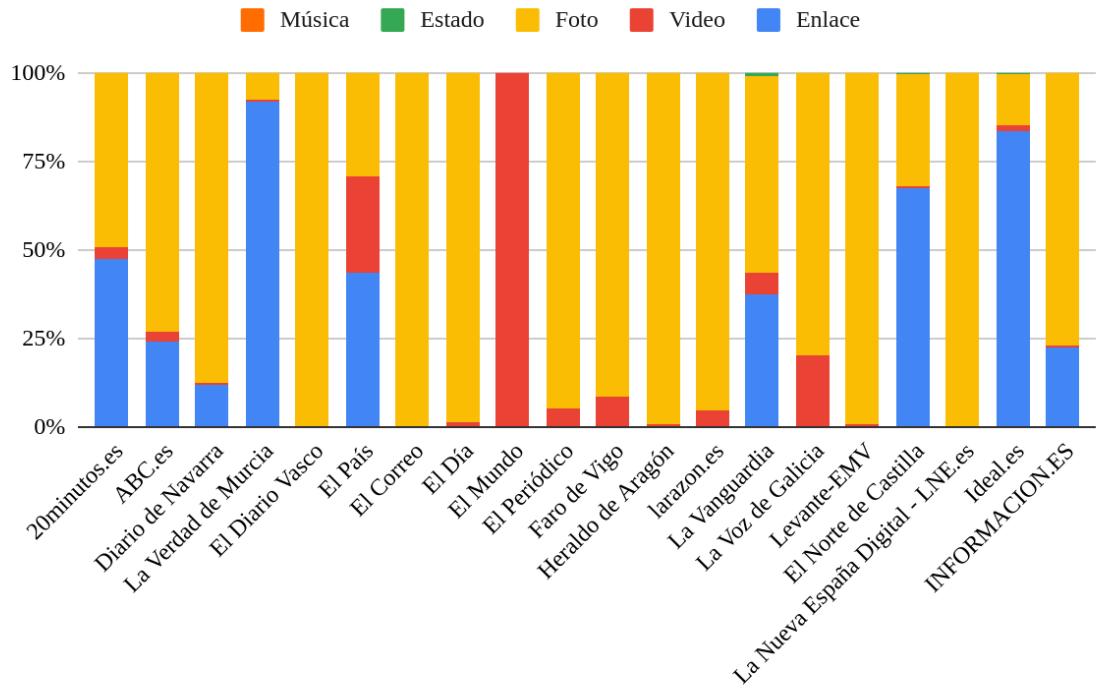
Como es apreciable, en general todos los periódicos tuvieron el mismo efecto en su incremento de interacciones durante el año 2016. No es observable una preferencia entre un periódico u otro distinta al propio tamaño o reconocimiento de cada uno, es decir, es normal

y esperado que *La Vanguardia* tenga muchas más interacciones que *El Correo*. Más allá del volumen de interacciones que tiene cada periódico de forma individual lo importante es observar el comportamiento uniforme de todos durante el año 2016.

Lo cierto es que no todos los periódicos presentan el mismo cambio en su parrilla de contenidos en el año 2016 tal como es observable en las siguientes figuras.

Figura 11

Participación de cada tipo de publicación en cada periódico durante el año 2015



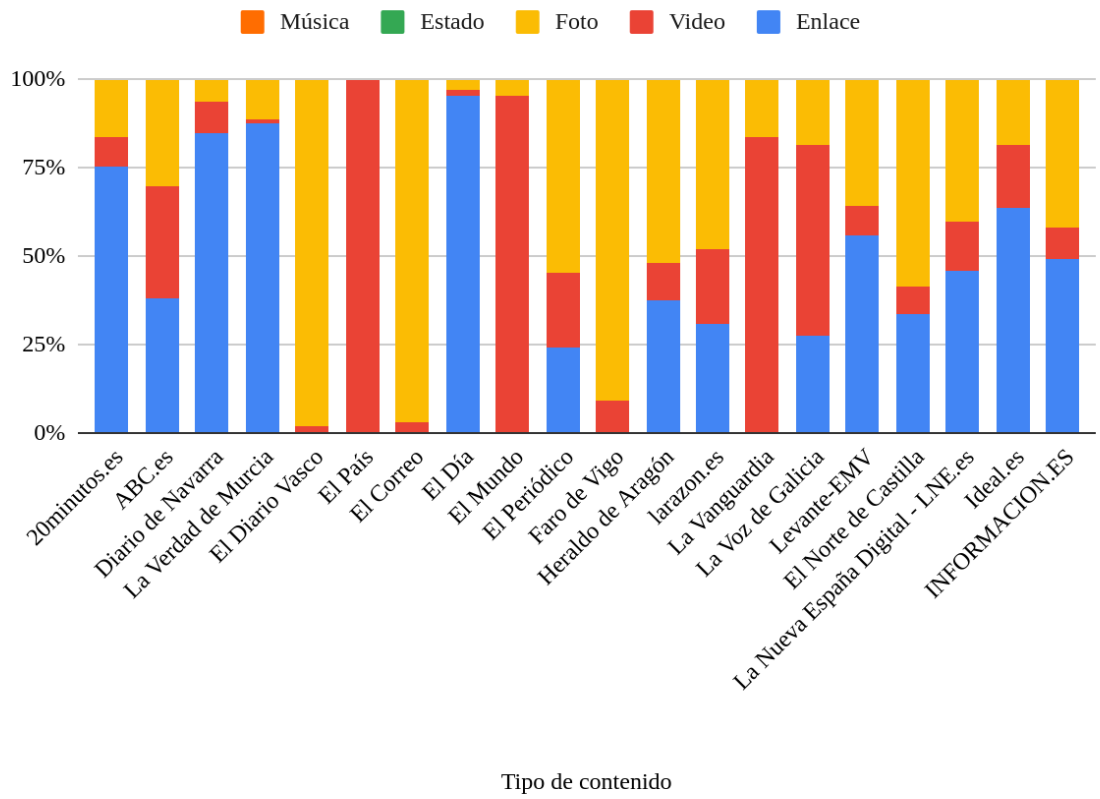
Fuente: Elaboración propia.

El año 2015 es un año en el que los periódicos en su gran mayoría publicaban contenidos de imagen -foto- y en donde pocos periódicos -*El País*, *El Mundo* y *La Voz de Galicia*- tenían una considerable participación de contenidos de video. Es un hecho de que en sí la gran mayoría de los contenidos publicados eran fotos o enlaces a las páginas de los

periódicos. Este es un aspecto que cambia de forma radical en el año 2016, aunque no de forma uniforme en cada uno de los periódicos.

Figura 12

Participación de cada tipo de publicación en cada periódico durante el año 2016



Fuente: Elaboración propia.

Para el año 2016 hay un cambio importante en la tendencia: se observa que los contenidos de foto caen de forma considerable en todos los periódicos salvo algunas excepciones como *El Diario Vasco* ya que la gran mayoría de periódicos empieza a publicar muchísimos más vídeos, es una tendencia que cambia por completo toda la tendencia de la gráfica, 2016 es

el año del vídeo. Es un cambio muy llamativo ya que tan solo un año antes (2015) había periódicos que no habían publicado ni un solo vídeo. Sea publicado muchos o pocos videos, el contenido de vídeos se hace extremadamente popular entre los distintos periódicos.

Un aspecto a tener en cuenta sobre estas gráficas es que debe recordarse que la información disponible no representa un muestreo uniforme o aleatorio de contenido, sino que incluye aquellos contenidos con mayor relevancia en cada uno de los perfiles. Con lo cual la visualización de un aumento en los contenidos de video no indica, necesariamente, un "aumento en las publicaciones de video" sino más bien una "mayor presencia de los contenidos de video entre las publicaciones más importantes de cada periódico". Aun así los datos disponibles tienen una correlación con el tipo de contenido más frecuente de cada periódico, lo cual implica que en todo caso estos datos sí reflejan un aumento de los contenidos de video.

Los contenidos "*click-bait*" siguen haciendo eco en 2017 cuando Facebook aumenta la precisión en la detección de estos contenidos mediante dos tipos de titulares: los que ocultan información o los que la exageran (Babu et al., 2017). El nuevo algoritmo persigue el contenido engañoso que usa técnicas propias de las noticias falsas como la exageración. Profundizando este estilo de detección de información malintencionada Facebook invisibiliza páginas malintencionadas que practican "*cloacking*". Esta técnica posibilita compartir contenido pornográfico, violento, de odio, etc. Gracias a la inteligencia artificial desde 2017 el algoritmo de Facebook elimina este tipo de contenidos (Leathern y Chang, 2017). Este es uno de los primeros casos en la historia de Facebook en donde utilizan la Inteligencia Artificial para filtrar contenidos, aunque sea de forma indirecta.

La revisión del contenido no está limitado al "*click-bait*" o al "*cloacking*", Facebook también aplica un algoritmo dedicado a detectar contenido de baja calidad en donde los usuarios tienen una mala experiencia con el fin de invisibilizarlo en el *News Feed* (Lin y Guo, 2017). De forma similar nuevas técnicas y algoritmos aparecen en 2017 para detectar y penalizar publicaciones que utilizan el "*engagement baiting*" (Silverman y Huang, 2017) -expresión traducible al acto de presionar al usuario para obtener una interacción, por ejemplo "si le das me gusta Microsoft donará 1€ a los niños de África"-.

Con esto se observa una clara tendencia: una invisibilización de lo que Facebook considera baja calidad. Esto a partir de la detección de prácticas malignas para los usuarios, aunque también aplican cambios que premian la "alta calidad". Por una parte, los videos que los usuarios reproducen durante más tiempo tienen mayor importancia en el *News Feed* (Bapna y Park, 2017) y, además, la relevancia aumenta aún más si el usuario trata de encontrarlo -por ejemplo, visitando la página que lo ha publicado- o si lo visualiza repetidamente (Facebook, 2017). Ahora no basta con "publicar videos", los videos también tienen que promover la permanencia de los usuarios en Facebook. Si los videos noticiosos no aumentan la permanencia, no tendrán mayor visibilidad en el *News Feed*.

Facebook cierra 2017 con cambios que afectan directamente a los periódicos como premiar los contenidos auténticos y medir señales de relevancia en tiempo real (Lada *et al*, 2017). Estos cambios deberían favorecer a los periódicos dado que su contenido es original y en algunos casos de importancia inmediata cuando hablamos de noticias de última hora. Estos cambios son benignos aunque no dejan de ser reglas a seguir para obtener visibilización en el *News Feed*. Consecuente con esto Facebook sigue convencido de que el *News Feed* es una herramienta de información y las noticias cuya fuente es de alta calidad obtienen una mayor exposición (Mosseri, 2017). Esto demuestra que Facebook conoce su poder de influencia sobre los usuarios al ser quien decide qué información distribuye, pero al mismo tiempo sabe que el contenido de calidad debe provenir de algún lugar.

También hay cambios menores que cambian el comportamiento del usuario, como la exposición de artículos relacionados (Su, 2017) y el botón de "*snooze*" -diseñado para que un usuario bajo su propia decisión no vea publicaciones de un usuario o página en particular durante 30 días- (Muraleedharan, 2017).

Mientras el concepto de "*fake news*" emerge a nivel global y empieza a ser una preocupación para gobiernos y medios de comunicación, Facebook aplica cambios clave al filtrar noticias: revisar la fiabilidad de la fuente, la importancia de la información a nivel local y verificar que sean verdaderamente informativas (Mosseri, 2018). Facebook empieza a preferir la información de relevancia local para el usuario -que sea importante en su ciudad, o país- en el *News Feed* (Hardiman y Brown 2018).

Dado que lo anterior no es suficiente para combatir la información falsa, Facebook ejecuta acciones que consideramos clave: promover programas de verificación de hechos en varios países, pruebas para conocer la veracidad de distintas fotos y videos, aplicación de nuevas técnicas para encontrar material duplicado (como el *Machine Learning*) y mejoras en las relaciones con la comunidad académica (Lyons, 2018). Este comportamiento de Facebook para combatir noticias falsas demuestra que requiere la colaboración activa de otras organizaciones especializadas. Es decir, Facebook no puede estar solo en esta lucha ya que es incapaz de lograr soluciones oportunas sin apoyo externo. Tanto es así que pese a que Facebook admite utilizar técnicas de *Machine Learning* para detectar noticias falsas también incorpora revisores internos y externos para realizar verificación de contenidos compartidos dentro de Facebook (Carden, 2018). Un algoritmo no es suficiente para detener la desinformación.

Para complementar lo anterior Facebook lanza nuevas herramientas para que cada usuario pueda encontrar más información sobre la fuente, el contexto de la noticia o los amigos involucrados en la difusión de esta de forma que un usuario pueda evaluar con mayor facilidad la veracidad de la información distribuida a través del *News Feed* (Hughes et al., 2018). Con estos cambios la reputación de una fuente de noticias es clave para la distribución del contenido en el *News Feed*.

Paradójicamente mientras hay tantos esfuerzos por filtrar *fake news*, también hay cambios que promueven la existencia de las cámaras de eco al visibilizar más las publicaciones considerando calidad y profundidad de la conexión entre usuarios. Ahora un usuario verá muchas más publicaciones de sus mejores amigos y menos de otros amigos con los que en raras ocasiones interactúa (Mosseri, 2018). Además, Facebook permite "silenciar" palabras clave, así que durante 2018 cada usuario puede ocultar durante 30 días todas las publicaciones en el *News Feed* con palabras claves concretas -por ejemplo PSOE o VOX- sin importar su fuente (Muraleedharan, 2018).

Para cerrar el estudio cronológico de *News Feed* tenemos el año 2019, un año profundamente marcado por la persecución de las *fake news* y de las organizaciones detrás de estas. Una de las acciones más frecuentes fue la expulsión de distintas entidades organizadas alrededor del mundo dedicadas a la desinformación en donde se encuentra una

gran cantidad de ejemplos como la eliminación de 364 páginas desinformativas originarias de: Rusia, Irán, Indonesia y más (incluyendo empresas como Twinmark Media Enterprises), Gleicher (2019). Este tipo de exclusiones en Facebook tienen antecedentes políticos desde finales de 2018 cuando Facebook elimina perfiles y páginas (incluyendo también perfiles de Instagram) pertenecientes a representantes y organizaciones militares de Myanmar, Slodkowski, A. (2018). Este tipo de situaciones ponen en evidencia la influencia ideológica y política de Facebook y cómo el *News Feed* es una herramienta para distribuir información con gran impacto social. Sin un filtro y control eficiente las *fake news* en el *News Feed* son un arma muy potente.

2019 también es el año en donde los resultados tecnológicos y humanos de Facebook para solucionar el problema de las *fake news* en el *News Feed* empiezan a incrementar. Encontramos ejemplos que menciona Woodford (2019) donde el uso de *Machine Learning* y revisores externos son clave para detener noticias falsas en la Unión Europea. Esto está conectado directamente con los esfuerzos de Facebook para restringir la exposición de las *fake news* en el *News Feed*: técnicas de *Machine Learning* para detectar contenido falso, revisores externos para verificar la información de dichos contenidos, exponer los artículos de los revisores como "artículos relacionados" en cada noticia verificada, tener un histórico de reputación para cada dominio y restringir aquellos que publican contenido malicioso constantemente, entre otras, Zigmond (2018). Incluso Silverman (2019) admite que Facebook se prepara para tener más alianzas colaborativas para detectar noticias falsas en el *News Feed*.

Facebook (2019) en su artículo "People, Publishers, the Community" admite saber que el estar en *News Feed* influye directamente sobre los contenidos que los usuarios producen y al mismo tiempo anunció medidas que precisamente buscan moldear lo que debería publicarse en el *News Feed*: invisibilizar contenidos que no suelen gustar a los usuarios o que no son originales, entregar experiencias personalizadas en el *News Feed* a cada usuario, mostrar comentarios que "generan mayor valor" al ocultar aquellos que son ofensivos, penalizar páginas o sitios web que suelen romper las normas de Facebook y perseguir la desinformación. Este es otro ejemplo claro de la manipulación mediante la invisibilización, Facebook sabe que tiene el poder de cambiar qué producen y publican los centros de

noticias y cambia el comportamiento de sus algoritmos para que el *News Feed* incluya lo que Facebook considera como adecuado sin que exista un consenso o un trabajo en equipo con todos los terceros involucrados.

Un aspecto llamativo de 2019 es la altísima perspectiva política que Facebook toma, el artículo "What Is Facebook Doing to Address the Challenges It Faces?", Facebook. (2019), deja claro cuáles son las medidas tomadas por Facebook para proteger las elecciones a nivel global -cosa que ejemplifica a lo largo de 2019 en múltiples artículos de su sala de prensa-, estas prácticas corresponden a las mencionadas anteriormente añadiendo que también que los anuncios publicitarios con fines políticos debe ser más transparentes. Facebook de forma explícita toma el control del discurso político, por lo menos desde el acto de moderar y filtrar la información. Esto lo posiciona políticamente y abre caminos a nuevas investigaciones sobre su influencia en la política exterior: ¿Ha influido Facebook sobre el comportamiento de los gobiernos? ¿El eliminar o invisibilizar páginas oficiales de gobiernos que no se alinean con la visión ética y moral de Facebook ha tenido influencia sobre las votaciones de un país?

A grandes rasgos los cambios mencionados en "People, Publishers, the Community" están reflejados en otros artículos como la actualización en el ranking de videos que explica Miller (2019), en donde los videos originales, con mayor repetición y mayor retención de usuarios son los que tendrán mayor visibilización en el *News Feed*.

Finalmente 2019 incluye otros cambios con índoles más "funcionales" como la nueva pestaña de noticias, Newton, C. (2019) -demostrando una vez más la importancia que tienen las noticias para Facebook-, también Rosen (2019) explicó las nuevas medidas más drásticas para impedir la distribución de contenido dañino a través de los videos en vivo además de nuevas investigaciones para comprender cómo se manipula el contenido para evadir los filtros de Facebook. También aplicaron encuestras para personalizar las publicaciones que cada usuario visualiza en el *News Feed*, Sethuraman *et al* (2019) y mayor información a los usuarios para explicar por qué ven determinados contenidos en Facebook Sethuraman (2019). 2019 es sin dudas un año muy político para Facebook en donde los cambios algorítmicos son la continuación de lo visto en años anteriores y con aplicaciones orientadas a "evitar acciones" que hagan daño a los sistemas democráticos en

distintas regiones. Esto eleva los niveles de responsabilidad y demuestra el poder de influencia que Facebook tiene a nivel global.

Conclusiones y futuras investigaciones

Una vez cartografiado cómo cambia el algoritmo de Facebook con los años se observan varias tendencias; En primer lugar y como se ha demostrado tras analizar una veintena de periódicos durante más de 10 años, los periodistas actualmente están casi más pendientes de cómo cambia el algoritmo de Facebook que de quién preside un gobierno. Lo primero es esencial para la supervivencia de los medios. Como ha sido expuesto, si Facebook decide incrementar la importancia de los vídeos, los medios producirán más vídeos y si promociona la interacción o los *hashtag* también lo harán los medios. Consideramos que uno de los aspectos más relevantes de este capítulo ha sido recuperar y enlazar toda la información sobre cómo cambia el algoritmo porque a Facebook le cuesta mucho hacerla pública. Ya desde 2013 un informe del Pew Research Center alertaba de "que el 47% de los usuarios de Facebook en Estados Unidos se enteran de las noticias a través de la red social (Mitchel et al., 2013)." Eso significa el 30% de los estadounidenses se enteran de las noticias no a través de los medios, sino de lo que Facebook distribuía en su *News Feed*. Si hasta hace unos años, la fórmula de la Coca-Cola era el secreto mejor guardado, ahora es el algoritmo de Facebook. Es cierto, siguiendo con la analogía, que un laboratorio químico podía analizar los principales ingredientes. Y eso es lo que ha pretendido este capítulo. No obstante, no todo era analizable y lo mismo sucede con los algoritmos de Facebook.

La segunda conclusión de este estudio es cómo Facebook está interesado en cuantificar las interacciones de sus usuarios para ofrecerles un mejor producto. Pero visto desde otro punto de vista, también les llega a conocer muy bien al lograr aunar información personal sensible, aspecto que permite operar a otras empresas con esa información como se demostró en el escándalo de Cambridge Analytica. Todo esto lo hace Facebook a través del *Social Graph*, usando la teoría matemática de grafos. Este conocimiento desemboca en un poder de influencia que se torna profundamente político en 2019, lo cual abre oportunidades de investigación sumamente necesarias. La influencia política de los

algoritmos de *News Feed* junto a las decisiones de Facebook pueden estar cambiando la política a nivel global.

La tercera es que muchos de los cambios han contribuido a una mayor extensión de las *fake news* porque Facebook confía en que el usuario es capaz de distinguir lo verdadero de lo falso. Pero se ha demostrado que no es así. Influyen desde el sesgo cognitivo hasta el nivel cultural. En los últimos años, Facebook ha cambiado el algoritmo para modificar esta tendencia, pero se ha evidenciado que tiene que depender de agentes externos -*fact checkers*- y que no es fácil resolver esto desde las matemáticas lo que no quiere decir que en el futuro sí pueda hacerse porque por ahí va mucha de la investigación actual

Con todo, los cambios en el algoritmo de Facebook condicionan las vidas de miles de millones de usuarios en todo el mundo, modifican estrategias de medios de comunicación o influyen en resultados electorales y formas de gobierno. Es importante que la academia profundice la relación entre política, *fake news* y algoritmos a través del poder de alcance que tiene Facebook. En este sentido, este es el primer trabajo recopilatorio que entrega una cartografía completa de *News Feed* demostrando cómo estos cambios son influyentes y cómo Facebook es consciente de que la manipulación mediante la invisibilización es una realidad.

GOOGLE Y LAS *FAKE NEWS*: MECANISMOS DEL BUSCADOR PARA DETECTAR INFORMACIÓN FALSA

Introducción

Google es el buscador más usado del mundo y sus resultados de búsqueda tienen una clara influencia sobre la población global. A igual que Meta, Google ha invertido deforma sustanciosa en desarrollar tecnologías y estándares técnicos que permitan facilitar la detección de *fake news* -o al menos la detección de contenido confiable-. En este artículo se evalúa y aúna toda la información disponible sobre los métodos y estándares usados por Google para dicho fin.

Las noticias falsas son una amenaza global que a través de la desinformación busca obtener beneficios políticos o comerciales. Estas necesitan una distribución escalable a través de una audiencia de gran tamaño utilizando cualquier tipo de medio, herramienta. Dado que los medios digitales son los más económicos y eficientes para lograr esto servicios como el buscador de Google son los preferidos para poder cumplir con sus objetivos. Sin una distribución las noticias falsas no suponen ninguna amenaza y por ello gran parte del debate se centra en este aspecto: cómo detener o como menos controlar la distribución de las *fake news*.

En este capítulo se describe cuáles son los protocolos implementados por Google para detectar el contenido desinformativo, esta es una distinción importante ya que Google aclara que sus esfuerzos no se dedican a la persecución y al control de las *fake news*, sino del contenido desinformativo -una diferenciación importante tal como está recalcado en el marco teórico-. Dado que la desinformación es uno de los objetivos principales de las *fake news* se toma en cuenta dichas acciones como mitigaciones contra estas.

En la teoría actual existen muchas referencias sobre cómo funciona el buscador de Google pese a que su algoritmo es un secreto muy bien guardado. Lo que no está registrado es sobre cómo Google puede estar usando sus algoritmos u otras técnicas para detectar el contenido desinformativo.

Lo cierto es que Google detecta este contenido y lo invisibiliza en los resultados de búsqueda posicionando este contenido en las últimas páginas de resultados. De allí que nos planteemos la pregunta: ¿Qué métodos, mecanismos o herramientas usa Google para actuar de esta forma?

Este capítulo es un trabajo descriptivo en donde se valora todo lo que Google hace y dice que hace para evitar la distribución de las *fake news* o la desinformación y cómo se cumple o no se cumple con este propósito desde un punto de vista técnico.

Como es visto en el desarrollo de este artículo una de las características del algoritmo de Google es que busca entidades informativas altamente confiables, a las cuales da mayor visibilidad y posicionamiento en su buscador, por ejemplo, para una búsqueda como "los teléfonos móviles causan cáncer" sitios altamente confiables como cancer.gov tendrán las primeras posiciones mientras que otras entidades desconocidas y sin respaldo científico oficial como cuidatehoy.com no tienen oportunidad de ocupar la primera página de resultados, esto es lo que apunta la teoría.

Con lo anterior es posible preguntarse cuál es el proceso que Google tiene para determinar cuáles entidades son más confiables que otras desde el punto de vista del algoritmo, ¿por qué *cancer.gov* debe tener mayor posicionamiento que *cuidatehoy.com*? ¿es un proceso matemático fiable? ¿Existen revisiones manuales? ¿Se trata de una selección arbitraria de información? De la misma forma es necesario entender cómo algunas entidades reconocidas como el canal de televisión Cuatro, la Cadena Ser, la Radio Nacional de España (RNE), etc. a través de su trabajo periodístico y de entretenimiento pueden propiciar la existencia de contenido "potencialmente desinformativo" causado por contenido de entretenimiento como "Espacio en Blanco" de la RNE, o un programa de esoterismo como Cuarto Milenio. ¿Google puede reconocer el propósito de este tipo de contenidos? ¿Los algoritmos involucrados pueden diferenciar los hechos verídicos del esoterismo y los contenidos de misterio? Esto solo demuestra que para evitar la desinformación y filtrar las *fake news* no es un trabajo simple que no se limita en detectar a las entidades confiables para darles mejores posiciones; al contrario, para Google surge el complejo reto de entender el propósito de cada contenido y posicionarlo en las búsquedas que sean más apropiadas según la intención del usuario. Todo esto demuestra lo complejo

que es abordar la desinformación sin caer en la censura. En cualquier caso, el objetivo de este artículo es describir qué está haciendo Google para evitar la desinformación y las *fake news*.

Existe además otro problema que hasta ahora no se había tenido en cuenta pero que en función de lo que decida Google será un antes y después en los medios públicos: existe un creciente consenso en que si un medio no es libre del poder (sus responsables son elegidos por los políticos como en España la Agencia Efe o RTVE o en China, la agencia de noticias Xinhua) no son fiables puesto que no son periodismo sino propaganda. Todo esto dependerá de cómo lo considere Google y esto es lo que se describe aquí.

Google es una de las compañías tecnológicas norteamericanas con la capacidad de buscar, procesar y distribuir información a una escala global, llegando a servir miles de millones de búsquedas al día[1], dentro de las cuales se incluyen noticias y material informativo de todo tipo. Esto es una realidad gracias a los cálculos minuciosos de los algoritmos de Google en donde analizan todo tipo de señales (o en otras palabras, factores de posicionamiento como la calidad del texto, el uso de palabras clave, la fecha de publicación de un contenido, etc.) para definir qué anuncios llegan a los usuarios, predecir qué videos desean ver en YouTube, qué correos electrónicos son los más importantes para cada persona y, entre otras cosas, qué información es visible en sus resultados de búsqueda. Si consideramos que según varios estudios las personas confían en sus búsquedas online para tomar decisiones[2], es crucial para compañías como Google encontrar métodos eficientes para detectar el contenido desinformativo y evitar que este llegue al usuario final.

En una era en donde la economía de la atención es determinante para dar poder de mercado e influencia a una compañía y en la que el poder de distribución de información es uno de los factores con mayor potencial de determinar lo que será la corriente política predominante la clave es el saber qué información es falsa, qué información es verídica, qué entidades, organizaciones o personas tratan de manipular la opinión pública y sobre todo, entender cómo los filtros algorítmicos modernos pueden beneficiar o perjudicar la conservación de una opinión pública consciente y bien informada. Esto se hace aún más

[1] https://blog.hubspot.com/marketing/google-search-statistics
[2] https://www.pewresearch.org/fact-tank/2020/03/05/most-americans-rely-on-their-own-research-to-make-big-decisions-and-that-often-means-online-searches/

importante al ser compañías privadas las que tienen la mayor capacidad de filtrar, a su antojo, la distribución de la información global: Facebook y Google como un gran "duopolio", pero entran también otras compañías como Yandex, Baidu, Naver, Microsoft, entre otras. Lo importante aquí es que la presencia de instituciones públicas no es más que una presión política sobre las decisiones de estas compañías privadas, es decir, ningún gobierno tiene poder de acción sobre el desarrollo y futuro de los algoritmos, al menos no en este momento.

La importancia de este capítulo recae en la verificación y mención de los distintos factores y métodos de revisión que Google utiliza para detectar la desinformación y, al mismo tiempo, seleccionar a las entidades más confiables. Esto es importante para todo investigador que desee analizar la distribución de las *fake news* y el funcionamiento de los distintos outlets de noticias desde la perspectiva de un motor de búsqueda como Google ya que un análisis de este tipo no sería posible sin conocer la información dispuesta aquí. Es clave para cualquier investigador profundizar sobre los aspectos aquí mencionados ya que la proliferación de las *fake news* no es solo un fenómeno social, sino tecnológico.

Metodología

El alcance de este capítulo es el de hallar y clasificar los factores clave utilizados por Google para detectar contenido desinformativo y determinar cuáles son sus acciones actuales para limitar el alcance de información malintencionada en sus productos. Usualmente los factores y señales de posicionamiento y de detección de contenido desinformativo no son listados ni compartidos públicamente por Google, esto supone una dificultad en la investigación y a su vez hace que el trabajo de exploración y documentación sea la primera etapa del proceso: es necesario leer cientos de contenidos, definiciones y opiniones expertas para poder filtrar y encontrar cuáles son realmente los elementos que evalúa Google para detectar el contenido desinformativo -aquí también recae el valor investigativo-. Para poder encontrar, validar y explicar el objeto de estudio, metodológicamente se siguen cinco pasos:

1. Documentar cómo funciona *Google Search*. Se ha recolectado toda la documentación pública disponible que explica cómo funciona el motor de búsqueda

de Google. El objetivo es entender las fases en las que *Google Search* puede detectar las noticias falsas y cómo la documentación pública respalda nuestros resultados de investigación.

2. Encontrar y escribir las definiciones de Google para el contenido desinformativo y conocer su posición frente a las *fake news*. La importancia de este paso en la investigación está en que la forma en la que Google detecte a las *fake news* dependerá de lo que la compañía defina como "contenido desinformativo o malicioso". Siendo estas definiciones la base fundamental sobre la cual pueden entenderse las señales utilizadas por Google para detectar *fake news*. De nada importa lo que el resto del mundo considere que son las *fake news*, sólo vale lo que Google considere en esta situación.
3. Explorar y cuantificar todos los posibles factores que evalúa Google para detectar el contenido desinformativo. Este proceso requiere de la lectura de cientos de documentos para contrastar y validar su contenido con las declaraciones públicas de Google y sus proyectos asociados al periodismo. En una segunda fase, se ha validado técnicamente distintos medios de comunicación para comprobar cómo aplican en su código y prácticas de redacción aspectos que Google evalúa para detectar contenido desinformativo.

Cómo funciona Google

Dentro del contexto del consumo digital de medios, entendemos que Google es una de las principales fuentes de tráfico para sitios de distintas categorías: e-commerce, instituciones, corporaciones, noticias y periódicos, enciclopedias, foros, etc. Esto parte de un proceso matemático e informático que ocurre en distintas escalas y ciclos. Existen distintas formas para las cuales se pueden interpretar todos estos procesos en su conjunto, pero una manera adecuada de resumirlos es hacer un análisis en tres pasos propuestos por Google (2019) en su documentación sobre "Cómo funciona la Búsqueda de Google": "Crawling", "Indexing" y "Serving".

Rastreo (Crawling)

El crawling es un proceso ejecutado por un "web crawler" o "spider", un software que de forma automática y sistemática navega a través de distintas páginas web, Merriam-Webster. (2019). Por tanto, se puede entender el "crawling" como el proceso de navegar distintas páginas web a través de los enlaces (hipervínculos) que llevan a otras páginas, un proceso constante cuyo fin es detectar nuevas URL, es un proceso de exploración que debe contar con un punto de partida.

Este proceso necesita de nuevas URL de forma constante, pero Google no es específico sobre cuáles son sus fuentes exactas de información para descubrir nuevas URL. Existen distintas hipótesis sobre las fuentes usadas por Google con el fin de iniciar procesos de Crawling, tales como libros, bases de datos públicas, contenido enviado por usuarios y más. Smarty, A. (2009) expone una serie de factores teóricos e hipotéticos mediante los cuales Google podría conocer nuevas URL para su proceso de crawling, tales como la manipulación de URLs existentes, hallazgo de enlaces dentro de formularios, enlaces hacia imágenes, entre otros. Entre líneas la existencia de estas hipótesis permite suponer que Google no depende únicamente de un proceso de crawling para encontrar nuevas URL, incluso puede ser factible la existencia de una intervención humana en este proceso.

Lyons, K. (2020) reportó la existencia de enlaces de invitación a grupos privados de WhatsApp siendo indexados por Google, información que fue revelada originalmente por un tuitero[3]. En principio si Google siguiera al pie de la letra el funcionamiento de lo que define un "crawler" no debería ser posible para este buscador encontrar invitaciones privadas a grupos de WhatsApp, especialmente si estas no se encuentran enlazadas en la red. Este tipo de situaciones solo validan la posibilidad de que hipótesis como "Google descubre URL's al interior de correos electrónicos como Gmail" tal como plantea Smarty sean reales. Esto expone posibles nuevos problemas de privacidad que deben ser abordados desde esta perspectiva.

La idea general que debe prevalecer del proceso de Crawling es que se trata de un proceso técnico destinado a descubrir la información disponible en Internet detectando nuevas URL

[3] Reporte original de los enlaces de invitación privada a grupos de WhatsApp: https://twitter.com/JordanWildon/status/1230829082662842369

a través de los enlaces existentes en la red y de otras posibles fuentes externas. Proceso en el cual probablemente Google descubre URL's por medio de fuentes que no admite públicamente.

Indexación (Indexing)

Una vez una URL es descubierta inicia el proceso de Indexing. Según Google (2019) en el documento "How Google Search Works" este proceso consta de un análisis del contenido cuyo fin de intentar comprender el propósito de una página. Esto involucra el uso de un "*parser*", un software que ejecuta un proceso llamado "*parsing*". Para "*parsing*" es posible encontrar distintos significados similares que cambian según su uso, Stanford NLP Group. (2019) habla de dos tipos de parser en el contexto del análisis semántico: 1) los de lenguaje natural que se encargan de analizar las estructuras gramaticales de una frase (del lenguaje común de los hablantes); y 2) los probabilísticos: que utilizan el conocimiento de análisis anteriores, basados en estadística avanzada, cuya intención es mejorar el análisis de nuevas frases que en teoría, aún no han sido pronunciadas. Dicho de otra manera, podemos entender al proceso de "*parsing*" como la descomposición de un texto para convertirlo en una estructura de datos que son utilizados para analizar el contenido de un discurso a partir de los resultados matemáticos. Es decir, el lenguaje verbal se traduce al matemático para que sea reconocido por los algoritmos. Por tanto el proceso de *parsing* en Google se entiende como una parte estructural esencial del proceso de indexación. Google participa activamente en el desarrollo de métodos y sistemas más eficientes de "*parsing*".

El parsing es un proceso fundamental para Google, por ello Petrov, S. (2016) anunció en el blog de Google AI el lanzamiento de SyntaxNet, un "parser" desarrollado por Google que utiliza *Machine Learning* implementado en TensorFlow (una plataforma de *Machine Learning* de código abierto). Google (2019) afirma que SyntaxNet es una tecnología utilizada en sus diferentes productos. Adicionalmente Chang (2017) afirma que Google reconoce información clave para analizar un contenido noticioso, como autor, las citas y las referencias a partir de un proceso de parsing. Estos resultados matemáticos que son producto del análisis del parser son los que se utilizarán en el último paso del funcionamiento de Google Search: el "Serving and ranking".

El hecho de que Google utilice *Machine Learning* para un proceso de parsing significa que la forma en la que analiza matemáticamente el contenido no es controlada en su totalidad por la intervención humana. Esto da lugar a que los algoritmos presenten parcialidad frente a su forma en la que interpretan la información. Por este motivo Google no cuenta con métodos exactos para determinar qué es cierto y qué no lo es, no puede depender fácilmente del *Machine Learning* para detectar una información que es falsa, después de todo el proceso de "*parsing*" está diseñado para entender la información, no para valorar su veracidad.

Publicación y posicionamiento (Serving and ranking)

Este último paso hace referencia a otros algoritmos que Google utiliza para determinar qué contenido del índice es el más indicado para responder a una búsqueda, según indica Google (2019) en el documento de ayuda sobre el funcionamiento de su buscador. Las señales evaluadas en el proceso de serving and ranking configuran el tipo de respuestas que puede encontrar una persona en París y otra en Medellín sobre una misma búsqueda. Así, por ejemplo, si busco un tecnicismo amplio como "cáncer" obtendré respuestas diferentes en las dos ciudades, independientemente del idioma. El proceso de ranking es el que jerarquiza los resultados de una búsqueda.

La forma en la que el algoritmo de búsqueda de Google funciona es la siguiente: tiene en cuenta más de 200 variables a la vez como: ¿Este contenido es noticioso?, ¿Es noticia local o internacional?, ¿Es reciente? Esto se visibiliza en gran material público de Google destinado a explicar con simplicidad el funcionamiento de su buscador[4].

Existen muchos autores y profesionales que publican hipótesis e ideas sobre qué factores exactos evalúa Google para posicionar un contenido. Acharya (2019) por ejemplo lista más de 200 factores incluyendo algunos como la longitud del contenido, la velocidad de carga de una página, la frecuencia con la que la información de un contenido es actualizada, enlaces externos, etc.

Lo cierto es que Google (2019) señala en su documento "How Google Search Works" que añadió solo en 2018 3234 mejoras en su buscador. Motivo por el cual las hipótesis

[4] https://www.youtube.com/watch?v=BNHR6IQJGZs

existentes sobre los factores de posicionamiento de Google son difíciles de validar y llevar un registro de todo lo que ocurre en el buscador no es posible desde un análisis externo. Aún si Google implantara un mecanismo serio y conciso para detectar las noticias falsas en su buscador, no se podría conocer con exactitud ni cuáles son esos factores ni cómo cambian a través del tiempo. Google es opaco en su funcionamiento, y por ende no se puede tener una confianza absoluta sobre sus métodos y esfuerzos aplicados en la detección de *fake news*. La única opción posible es detectar y archivar los factores que hipotéticamente deberían afectar la detección y posicionamiento en el buscador de toda información dañina y correlacionar un análisis matemático a estos factores.

Google y las noticias falsas

Como ya es mencionado en el marco teórico y estado de la cuestión las *fake news* a pesar de ser un término ampliamente utilizado su definición a la fecha no es unánime. Gelfert (2018) define "noticia falsa" como la presentación deliberada de afirmaciones falsas o engañosas como si fuesen noticias. En este sentido, estarían diseñadas para ser imprecisas y desinformativas. El matiz de diseño intencionado es fundamental, frente a una información que pueda ser falsa de forma accidental. Esto plantea problemas importantes ya que las *fake news* pueden provenir del poder político o económico, de instituciones oficiales o privadas. Por ejemplo, toda la información política institucional que sea diseñada como propaganda puede ser engañosa. ¿Todo lo que afirma Trump es mentira? ¿Debe Google determinar qué es mentira y qué no de lo que declara Trump? ¿Todo lo que sale de la Casa Blanca es "*fake news*"? También medios de comunicación que no sean imparciales por un posible conflicto de interés con sus financiadores podrían incurrir en estas prácticas.

En este sentido, si Google quisiera detectar las noticias falsas debe valorar si la información que proporcionan instituciones públicas, gubernamentales o medios de comunicación vinculados a gobiernos puede ser considerada como información falsa. Es un debate que está abierto. En el caso que de Google lo considere de esta manera, como puede estar pasando con la información china procedente de la agencia china Xinhua (respecto al coronavirus, por ejemplo), habría que valorar qué pasaría con los medios de comunicación españoles como Agencia Efe o RTVE o todos los autonómicos públicos cuyos responsables

son puestos por los gobiernos. Google, de momento, no tiene interés en involucrarse con este problema. Si lo hiciera, acabaría con la noción de medios públicos de los países mediterráneos. Si no lo hace, no es efectivo en la detección de toda la información que es nociva para los ciudadanos. Esto sin contar la complejidad que supone el planteamiento de modelos matemáticos y escalables que permitan revisar estos factores en países enteros.

Esta lógica explica por qué en la documentación filtrada por Project Veritas. (2019), Google no se refiere directamente a las "noticias falsas" o "*fake news*" en sus políticas y comunicación interna, sino a la información mal intencionada o a la desinformación ("misinformation" en inglés). Google aclara esto debido a que el hablar de *fake news* están obligados como empresa a definir un concepto como "verdad" a nivel internacional. Esto sería un imposible porque implica toda clase de cuestiones pragmáticas y filosóficas; ¿Existe la verdad? Excepto en las ciencias naturales que usan un método científico muy estricto, ¿qué es la verdad? Una noticia sobre el Camino de Santiago puede tener un aspecto de verdad -el número de peregrinos de un día concreto- y otro que no lo es ¿realmente está Santiago enterrado ahí? Google no puede desmentir esta información -por ejemplo para no desacreditar creencias religiosas- y en lugar de hablar de *fake news* menciona al contenido tergiversado y el discurso del odio, lo que es un acercamiento más pragmático y seguro.

Lo anterior obliga a Google a depender en mayor medida de la legitimidad de la fuente para determinar si un contenido es desinformativo, no del contenido. Es decir, para un contenido falso como es la astrología, Google tiende a dar mayor importancia a la fuente. Si busco el horóscopo "sagitario" en las primeras posiciones saldrá el horóscopo que publican cabeceras periodísticas solventes, pero no cuestiona que esa información es totalmente falsa porque la astrología es una pseudociencia con intereses timadores. Esto demuestra la otra cara de Google como buscador: por encima de la verdad su principal interés es mostrar al usuario lo que está buscando, pero para evitar la exposición de contenido dañino tiende a posicionar mejor a una fuente reconocida.

Por ende, el fin último de la investigación será exponer los factores evaluados por Google para detectar lo que ellos consideran como información dañina para sus usuarios, todo enmarcado en las características y definiciones descritas por Google en el documento expuesto por Project Veritas.

Las señales utilizadas para detectar las noticias falsas

Detectar una *Fake New* es una tarea compleja para cualquier algoritmo por lo difícil que puede ser, incluso para un ser humano, determinar qué es cierto y qué no lo es. Aunque es cierto que existen estudios en donde con distintos modelos y técnicas es posible detectar con cierta precisión que un contenido es una *Fake New*, por ejemplo la detección de *fake news* a través de un trabajo público de todos los usuarios de Tchakounté et al (2020) o el sistema de detección de noticias falsas a través de *Deep Learning* de Li et al (2021). No obstante estas técnicas en general se basan en técnicas cuantitativas o de participación ciudadana que no están diseñadas para determinar qué es cierto y qué no lo es. Es decir, son modelos que pueden detectar el contenido desinformativo, pero no decir por qué es desinformativo.

Con Google hay un escenario similar: es muy probable que en su intento por detectar el contenido desinformativo deba recurrir a modelos matemáticos y técnicas algorítmicas avanzadas para determinar qué es y qué no es desinformativo, teniendo en cuenta que esto es un trabajo que requiere siempre de un feedback externo para su constante perfeccionamiento -sea de un ser humano, organización o fuente que confirme al sistema que en efecto ha dado con información falsa-. Por este motivo las fuentes oficiales, la verificación técnica en portales noticiosos y la lectura entre líneas de comunicados oficiales son las principales herramientas para exponer qué es un factor potencial para Google para detectar el contenido dañino y desinformativo.

Señales provenientes de datos estructurados

Para cualquier motor de búsqueda es necesario dividir la información en estructuras de datos que sean lógicas para su procesamiento. Tarea que al no ser simple y no tener resultados perfectos llevó a Google, junto a Microsoft, Yahoo y Yandex a fundar Schema.org, un proyecto que desde 2015 toma decisiones sobre la representación de datos estructurados en internet, Schema. (2019).

Los datos estructurados según la definición de Webopedia. (2020) corresponden a cualquier información alojada en un campo fijo dentro de un archivo. Esto quiere decir que todos los sitios web que incluyen datos estructurados y que además siguen los lineamientos de

Schema.org indicarán siempre de una forma estandarizada la información clave del contenido.

Entre más webs tengan datos estructurados, más fácil será para un motor de búsqueda el ejecutar un proceso de parsing limpio y eficaz para extraer información clave en un análisis del contenido. Por ejemplo, conocer el autor, fecha de publicación y organización de un artículo noticioso sin tener que intuir cómo se encuentra esta información dentro del contenido en sí (proceso que en la práctica sería más complejo e impreciso). Por ejemplo, si Google intenta obtener el autor de un artículo de *The Guardian* y también de un artículo en *CNN* y ninguno de estos dos sitios web ofrece su información con datos estructurados, Google no tendría un método exacto para saber dónde *The Guardian* escribe la información de sus autores, que además será en un diseño, lugar y forma completamente diferente al de *CNN* (y de cualquier otro portal de noticias). Pero como *The Guardian* y *CNN* tienen una implementación estándar de datos estructurados de Schema Google no tiene que "intuir" dónde está la información del autor, siempre verá en el código fuente de ambas webs el nombre del autor, en el mismo sitio, en el mismo formato. Gracias a Schema Google puede entender de forma inequívoca cuál es el autor en cualquier contenido de cualquier web.

Google (2020) ofrece una lista extensa de datos estructurados, para los cuales en el ámbito noticioso haremos hincapié en algunos de ellos: artículos, multimedia, transmisión en vivo y verificación de datos.

Tabla 1

Resumen y definiciones de objetos de datos estructurados

		Datos estructurados		
Objeto	**Definición**	**Artículo**	**Transmisión en vivo**	**Verificación de datos**
		Schema hecho para artículos ya sean artículos noticiosos, deportivos, de blog u otros.	Schema diseñado para indicar si un video presente en una web es una transmisión en vivo.	Schema que contiene información estructurada sobre la verificación de datos o "fact check".

		Datos estructurados		
Objeto	**Definición**	**Artículo**	**Transmisión en vivo**	**Verificación de datos**
Autor	Es el autor del contenido (ej. un artículo) y puede definirse como una persona o una organización	Incluido	No aplica	Incluido, como autor se entiende a quien publica el artículo de verificación de datos. También incluye al autor de la declaración que es evaluada en la verificación de datos.
Fecha de publicación	Fecha en la que el contenido es publicado.	Incluido	Incluido, aunque se refiere específicamente al momento en el que inicia la transmisión en vivo. También tiene valores similares como "uploadDate" que corresponde al momento en el que se publica el video por primera vez, y también "publication", el cual indica cuándo será transmitido en vivo el video.	Incluido, contemplando la publicación tanto del artículo de revisión como de la declaración que se evalúa.
Encabezado o "headline"	Es el titular de un artículo	Incluido	Incluye una propiedad similar llamada "name" que corresponde al título del video.	No aplica
Imagen	Imagen que representa y pertenece a un artículo o contenido.	Incluido	Incluye un objeto similar denominado "thumbnail" o imagen miniatura.	No aplica
Editor del artículo o "publisher"	El editor del artículo o del contenido. Debe ser una organización.	Incluido	No aplica	No se incluye ya que el "publisher" equivale al autor en este tipo de contenidos.
Fecha de modificación	Fecha y hora de la última edición del artículo o contenido.	Incluido	No aplica	No aplica
Descripción	Breve descripción del artículo o contenido	Incluido	Incluido	No aplica
Principal entidad de la página o "mailEntityOfPage"	Es la URL canónica del artículo. Debe especificarse si el artículo o contenido corresponde al tema principal dentro de una página.	Incluido	Incluye dos objetos similares, uno llamado "contentURL" que dirige al archivo multimedia de video, y otro llamado "embedURL" que incluye un reproductor para el video.	Incluye un objeto similar llamado "url" que corresponde a la página que contiene la revisión completa.
Logo del editor	El logotipo del editor del artículo o contenido	Incluido	No aplica	No aplica
Duración	Indica cuánto dura un video	No aplica	Incluido	No aplica

		Datos estructurados		
Objeto	**Definición**	**Artículo**	**Transmisión en vivo**	**Verificación de datos**
Expiración	Especifica cuándo dejará de estar disponible un video.	No aplica	Incluido	No aplica
Evento de broadcast	Grupo de objetos que especifican la fecha de finalización de la transmisión, si el video se encuentra en vivo y la fecha de inicio de la transmisión.	No aplica	Incluido	No aplica
Revisión de reclamo o "claimReviewed"	Se trata de un resumen la declaración revisada.	No aplica	No aplica	Incluido
Rating de revisión	Es la valoración de la declaración revisada, la cual se cuantifica desde 1 (falso) hasta 5 (veraz). Los valores intermedios se representan como: 2 (falso en su mayor parte), 3 (parcialmente veraz), 4 (veraz en su mayor parte). Se añade una cuantificación especial de "-1" que corresponde a "difícil de clasificar".	No aplica	No aplica	Incluido

Fuente: Tabla de elaboración propia con información de Google https://developers.google.com/search/reference/overview, se han resumido los elementos necesarios en cada uno de los Schema para una interpretación más fácil.

La información reunida en estas tablas es importante debido a que destaca cada elemento necesario según el tipo de contenido al cual se aplica un dato estructurado -*Schema*-. Aunque en schema.org existe soporte para un mayor número de datos estructurados, estos tres son los que se asocian a un contenido periodístico, y por tanto los que deben tener especial atención a la hora de detectar noticias falsas. La información que Google interpreta dentro de los datos estructurados es importante para el proceso de publicación y posicionamiento. Esto hace lógico que los factores y señales utilizados para detectar noticias falsas utilicen parte de los elementos listados en la tabla, en especial si un proceso de "parsing" tiende a ser impreciso. Google necesita de los datos estructurados para

conocer con precisión el contenido de una web, y por ello en su lectura de los datos estructurados de Schema busca elementos clave de veracidad de la información: autor, fecha, fotografías, verificación de datos, titulares, etc. La tabla anterior otorga pistas fundamentales sobre lo que Google puede leer y verificar sobre un contenido con precisión.

Entidades

Para una mayor precisión en la interpretación semántica Google extrae y analiza "entidades", las cuales están definidas en la patente de Google US9477759B2 publicada por Keysar, D., & Shmiel, T. (2013) como una cosa o concepto singular, único y distinguible de otros. Por ejemplo, un lugar (ej. Madrid), una persona (ej. Juan Manuel Santos, expresidente de Colombia) o un sentimiento (ej. ansiedad). Esto es importante, porque los datos estructurados permiten a Google reconocer sobre qué trata un contenido exactamente y valorar la calidad, idoneidad y propósito de los contenidos en relación a las entidades que contienen. Por ejemplo, no tiene el mismo peso un artículo periodístico escrito por *The Guardian*, periódico que es una entidad reconocida y antigua con décadas de historia en comparación a una entidad desconocida y recién constituida como puede ser un medio independiente con escasos meses de existencia. Lo mismo puede ocurrir con un autor, si Google llega a reconocer a un autor como alguien mucho más reconocido y confiable sus contenidos deberían tener un mejor posicionamiento. Esto de forma implícita permite asumir que la exposición en Google es proporcional al reconocimiento que tienen las entidades involucradas: autores y editoriales principalmente. Pero a su vez, las entidades reconocidas dentro del contenido en sí mismo también son valoradas, es evidente que una noticia sobre una entidad sensible como "Coronavirus" requiere mayor revisión y verificación que las noticias asociadas a otras entidades de menor impacto social, económico o político.

Lo verdaderamente importante de las entidades como señal de evaluación es que ayudan a expandir el entendimiento que se tiene de las búsquedas, las cuales no se limitan al uso de palabras clave, sino que ahora Google considera conceptos e ideas generales que tienen el potencial de ser un factor en la detección de noticias falsas.

Esfuerzos de Google para combatir las noticias falsas

El análisis de las señales utilizadas por Google para detectar noticias falsas no se limita a las capacidades técnicas de su buscador, también debe incluirse un análisis de sus esfuerzos y aportes en proyectos externos. Uno de ellos es el "Google News Initiative", un proyecto destinado a ayudar al periodismo en la era digital, Google. (2018). Este proyecto incluye toda clase de iniciativas y entre ellas es *The Trust Project* el más importante en la detección de noticias falsas.

The Trust Project

The Trust Project -también conocido como TTP- es fundado por Google y otras compañías para trabajar en conjunto con más de 75 organizaciones periodísticas de todo el mundo, Chang (2017). El objetivo de TTP es generar un indicador de confianza para las noticias publicadas para que sea posible distinguir la calidad y el propósito de un contenido (si es comercial, de baja calidad, de alta calidad, si se trata de un contenido de opinión, etc.)

TTP tiene ocho indicadores de confianza: mejores prácticas, experiencia del autor, tipo de trabajo, citaciones y referencias, métodos, fuentes locales, diversidad de voces y finalmente feedback accionable. En la página principal de *The Trust Project* (2019) se definen a estos ocho indicadores denominados "The Trust Indicators" (los indicadores de confianza) como una divulgación estandarizada sobre ética, justicia y precisión de los medios de comunicación siendo los primeros en otorgar a buscadores y redes sociales estándares técnicos para evaluar la calidad de una noticia. Según TTP son utilizados por compañías asociadas como Facebook, Google y Bing. Esto nos da la certeza de que la información expuesta y evaluada por TTP es utilizada por Google para detectar contenido malicioso o desinformativo.

El principal problema de los indicadores de confianza es que dan lugar a muchas ambigüedades y dependen de una interpretación subjetiva -Por ejemplo, no está definido con exactitud cómo se determina que una referencia es de calidad-. Además de ello por diseño estos indicadores no otorgan métodos que impidan la falsificación de información o la detección de información falsa, por ejemplo: no hay en la teoría nada que impida que un medio reconocido publique información falsa y que al mismo tiempo esté certificado por

TTP. Por esto es posible debatir sobre su eficiencia y utilidad en el uso práctico de esta herramienta.

Aunque su eficacia sea cuestionable eso no significa que un indicador de confianza no pueda ser analizado por Google mediante los procesos de parsing e indexación, en procesos de análisis con el uso de *Machine Learning*, Inteligencia Artificial o incluso en revisiones manuales. Es muy probable que estos indicadores sean un punto de evaluación para el corpus de noticias de Google.

Best practices

Este indicador provee información sobre las buenas prácticas necesarias para que pueda ser distinguido el periodismo de otro tipo de contenidos, Pensiero et al. (2019). Por ejemplo, distinguir una noticia de un artículo de opinión. El documento oficial incluye distintos factores a evaluar para la construcción de este tanto a nivel de sitio web como a nivel de un contenido individual.

El indicador de "*Best practices*" tal como lo describe Pensiero et al. (2019) recolecta la información necesaria para visibilizar las políticas de ética, correcciones, clarificaciones, privacidad y diversidad. También es un indicador que busca fomentar la exposición de correcciones y clarificaciones sobre el contenido publicado así como la declaración oportuna de quiénes son propietarios y/o financian al medio que publica un contenido. Por otra parte este indicador expone la necesidad de aclarar las garantías de independencia periodística y al equipo editorial junto a la fecha en la que el medio de comunicación es fundado. Incluye además las prioridades de cobertura periodística y sus formas de contacto, aclaraciones sobre sus procesos de verificación y la posibilidad de tener retroalimentación accionable y por último busca una referencia de participación a TTP solo si el medio de comunicación ha sido aceptado dentro del proyecto.

Este indicador de confianza no otorga herramientas claras para que un algoritmo pueda verificar si un contenido es engañoso o no; simplemente se limita a facilitar mediante un estándar la exposición de políticas y prácticas que cualquier medio de comunicación maduro debe de cumplir. El punto más llamativo de este indicador es la posibilidad que tiene un medio de comunicación de indicar si ha sido aceptado por TTP, lo cual lleva de

nuevo a un debate profundo sobre las noticias falsas: ¿quién puede verificar que los medios aceptados por TTP exponen la verdad? ¿Es posible que un medio aceptado por TTP publique *fake news* en el futuro? ¿Puede TTP detectar el incumplimiento de sus buenas prácticas y expulsar a un medio que ya fue aceptado? No existe una documentación que responda con exactitud a estas preguntas.

El enfoque del indicador "*Best practices*" es calificar la integridad de cualquier medio de comunicación según su cumplimiento de las prácticas "necesarias" para exponer información veraz que pueda ser corregida y que además sea independiente. Esto en la teoría permite evitar que un medio de comunicación grande o pequeño caiga en la divulgación de noticias falsas, pero no es una garantía de ello.

Esto hace de este indicador una señal relativamente débil ya que no otorga garantías de cumplimiento más allá de una verificación manual, es decir, esta información no reemplaza el trabajo humano detrás de la verificación de las noticias falsas, su única posible función es facilitar información clave para la verificación manual a aquellas personas que tienen la responsabilidad de hacerlo.

Aunque la verificación de estos datos sea manual (la lectura, análisis y verificación de cumplimiento de estas prácticas), evidenciando que los medios están exponiendo esta información mediante datos estructurados de schema.org

```
<script type="application/ld+json">
{
"@context": "http://schema.org",
"@type": "NewsMediaOrganization",
"name": "BBC News",
"ethicsPolicy": "http://www.bbc.co.uk/editorialguidelines/guidelines",
"masthead": "http://www.bbc.co.uk/news/help-41670344",
"missionCoveragePrioritiesPolicy":
"http://www.bbc.co.uk/news/help-41670342#missionstatement",
"foundingDate": "1922-10-18",
"diversityPolicy":
```

```
"http://www.bbc.co.uk/news/help-41670342#diversitypolicy",
"correctionsPolicy":
"http://www.bbc.co.uk/editorialguidelines/guidelines/accuracy",
"verificationFactCheckingPolicy":
"http://www.bbc.co.uk/editorialguidelines/guidelines/accuracy",
"unnamedSourcesPolicy":
"http://www.bbc.co.uk/editorialguidelines/guidelines/fairness/anonymity",
"actionableFeedbackPolicy":
"http://www.bbc.co.uk/editorialguidelines/guidelines/accountability/feedbac
k-and-complaints",
"ownershipFundingInfo":
"http://www.bbc.co.uk/news/help-41670342#ownership",
"diversityStaffingReport":
"http://www.bbc.co.uk/diversity/strategy/eir-2017",
"contactPoint" : [{
      "@type" : "ContactPoint",
      "contactType" : "Newsroom Contact",
      "email" : "haveyoursay@bbc.co.uk",
      "url" : "http://www.bbc.co.uk/contact#news"
  },
  {
      "@type" : "ContactPoint",
      "contactType" : "Public Engagement",
      "email" : "haveyoursay@bbc.co.uk",
      "url" : "http://www.bbc.co.uk/news/have_your_say"
  }]
}
</script>
```

Fuente: Código fuente extraído de https://www.bbc.com/news/help-41670342

Este segmento de código encontrado en bbc.com es información en formato *JSON-LD* que cumple con los estándares de schema.org. En este código podemos ver cómo los requerimientos del indicador "Best practices" están listados y enlazados. La forma en la que la información se estructura y expone tuvo que ser propuesta por TTP, además es esta la información que Google, Facebook y Bing interpretan en el código fuente en una web noticiosa para entender si cumplen o no con las mejores prácticas de integridad periodística, o al menos saber dónde están para una futura evaluación manual. TTP está transformando la forma en la que las webs noticiosas se construyen.

Journalist Expertise

Este indicador reúne información sobre la pericia de los periodistas. En el documento de Chance et al. (2018) está especificada la información relevante sobre los autores: nombre, residencia, idiomas que habla, temas, localidades y demografías en las que es experto, biografía, detalles de contacto, identificadores (un perfil en Twitter, por ejemplo), afiliaciones (si es empleado, freelance, independiente, etc.), títulos y roles (si es reportero médico, tecnológico, jefe editorial, entre otros) y, además, especificar en qué página se existe un archivo de sus artículos y biografía. Adicionalmente, y sin ser obligatorio, parte de la información recomendada para incluir en la información de autor se encuentra su fotografía, honores, reconocimientos, y membresías profesionales tales como federaciones, asociaciones, fundaciones, etc.

Todos estos indicadores permiten a cualquiera, tanto humanos como robots, identificar con precisión cualquier autor con el fin de cualificar la calidad de un contenido periodístico. Aunque es obvio que la experiencia y reconocimiento de un autor es un aval inequívoco que ayuda a respaldar la veracidad, es cierto que esta información es fácil de falsificar: después de todo la existencia de esta información en una página web no garantiza ni que el artículo sea verídico ni que el autor haya escrito realmente determinado contenido. A decir verdad, este indicador de TTP puede ser rellenado por cualquiera y no facilita una "garantía técnica" de que tanto la autoría como la información del autor sean ciertas. Lo anterior no quiere decir que esta información no sea útil para el análisis de una noticia y para evitar la existencia de un contenido desinformativo, pero su utilidad sigue dependiendo de la confianza que se deposita en un medio de comunicación específico. Esto obliga a Google y a cualquier otro que quiera analizar la veracidad de la información a depender de los medios más solventes y no íntegramente sobre los indicadores de TTP.

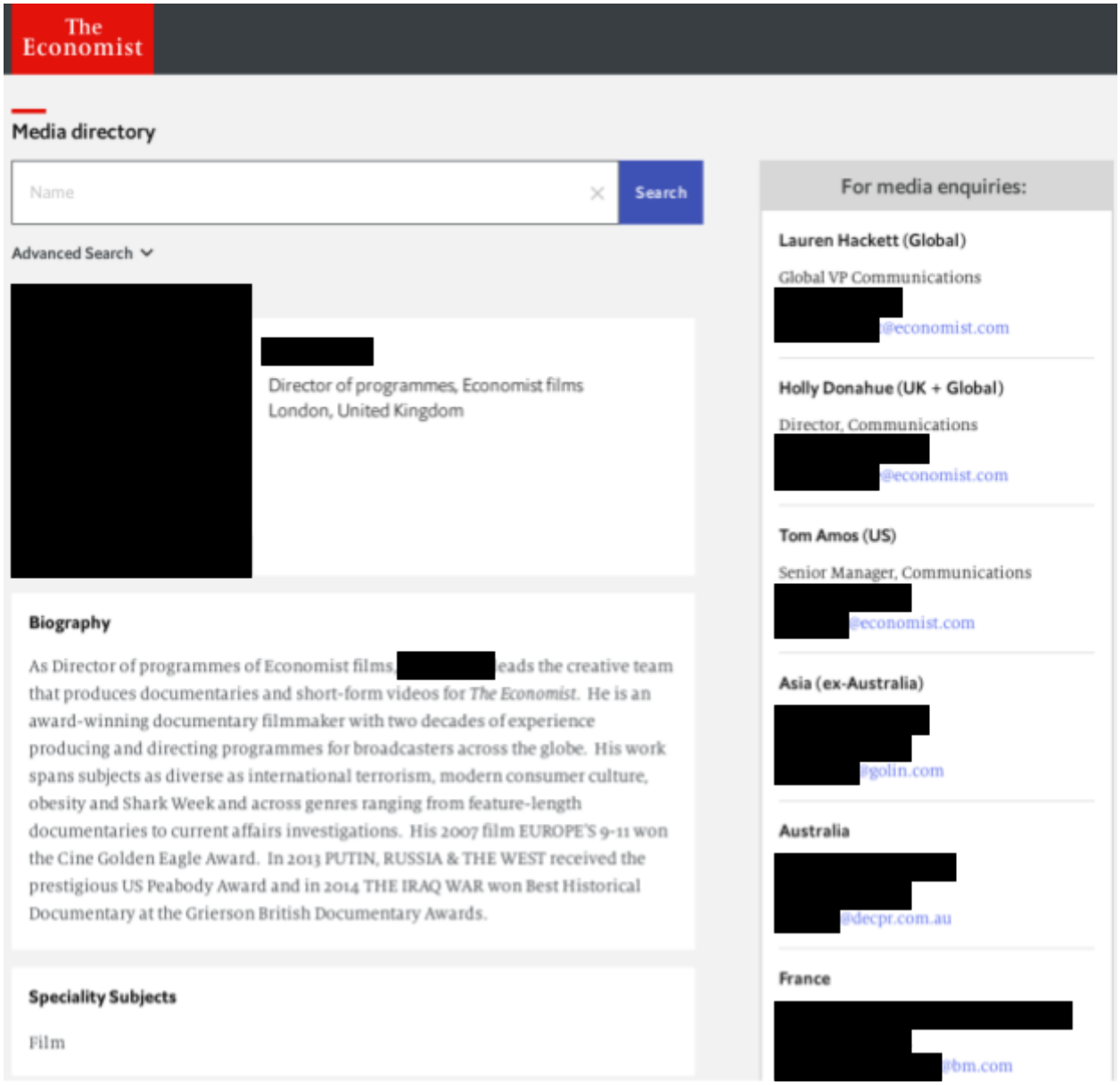
Uno de los aspectos más interesantes sobre este indicador de confianza es que permite mediante el análisis del enlazado, palabras clave y perfiles sociales el hallazgo y reconocimiento de un autor a través de Internet. Esto permite a través del análisis de entidades asociar artículos de distintos medios a un único autor. Con esto se logra atribuir cada artículo a un autor para construir su "reputación digital" además de calificar cada artículo según la reputación de su autor. Esto permite detectar contenido potencialmente

dañino según el autor, pero no otorga ninguna garantía que impida falsificar la autoría de un artículo. Este indicador de confianza necesita de un fuerte apoyo tecnológico para que sea útil y su contenido no debería ser aceptado ciegamente.

No todos los periódicos siguen las recomendaciones de TTP al momento de escribir este artículo, editoriales como *El País*, *BBC* o *The Wall Street Journal* no cumplen con todos los requisitos listados (pese a ser alguno de estos listados como los primeros portales en aplicar las recomendaciones de TTP). Aunque existen algunos otros como *The Economist* que siguen estas recomendaciones es cierto que dicha información no está contenida en datos estructurados, lo que dificulta su análisis.

Figura 1

Página con biografía de un autor en The Economist



Fuente: Imagen extraída de economist.com, toda la información personal o pública referente a la fotografía del autor o de contacto ha sido removida de forma intencional.

Uno de los ejemplos más destacados se encuentra en la sección de autores de *economist.com*, en donde observamos cómo distintos aspectos requeridos por TTP son especificados dentro de la página, pero no son especificados mediante datos estructurados. Dado que TTP permite a *The Economist* exponer en su web el logo de la organización,

significa que los datos estructurados no son obligatorios para cumplir con las recomendaciones de “Journalist Expertise”.

Otros portales exponen información clave del autor en formato Schema:


```
{
        "@type": "Person",
        "@id": "193937",
        "name": "Thore Haugstad",
        "jobTitle": "Freelance writer",
        "url": "https://www.fourfourtwo.com/users/thore-haugstad",
        "sameAs": "https://twitter.com/Haugstad1006",
        "description": "You might call Thore Haugstad pan-European: born in Norway but now based in Madrid, he possesses\u00a0an enquiring mind which roves across the continent for the finest stories. He covers every blade of grass on the pitch of football writing: from tactics to trends, stats to opinions, retro reminiscences to futurology.\u00a0",
        "workLocation": "Madrid",
        "knowsLanguage": "en, nb, nn, es",
        "knowsAbout": "Spanish football Tactics Analysis"
}
```


Fuente: Código fuente extraído de https://www.fourfourtwo.com/users/thore-haugstad

En este ejemplo la información es estructurada en formato *JSON-LD*, al igual que el indicador “Best Practices”. Los datos estructurados tienen como finalidad la exposición organizada de la información presente en la página de autor. El hecho de que estos datos se encuentren en un formato estructurado demuestra el interés de los portales noticiosos de facilitar la interpretación y análisis a compañías como Google y Facebook.

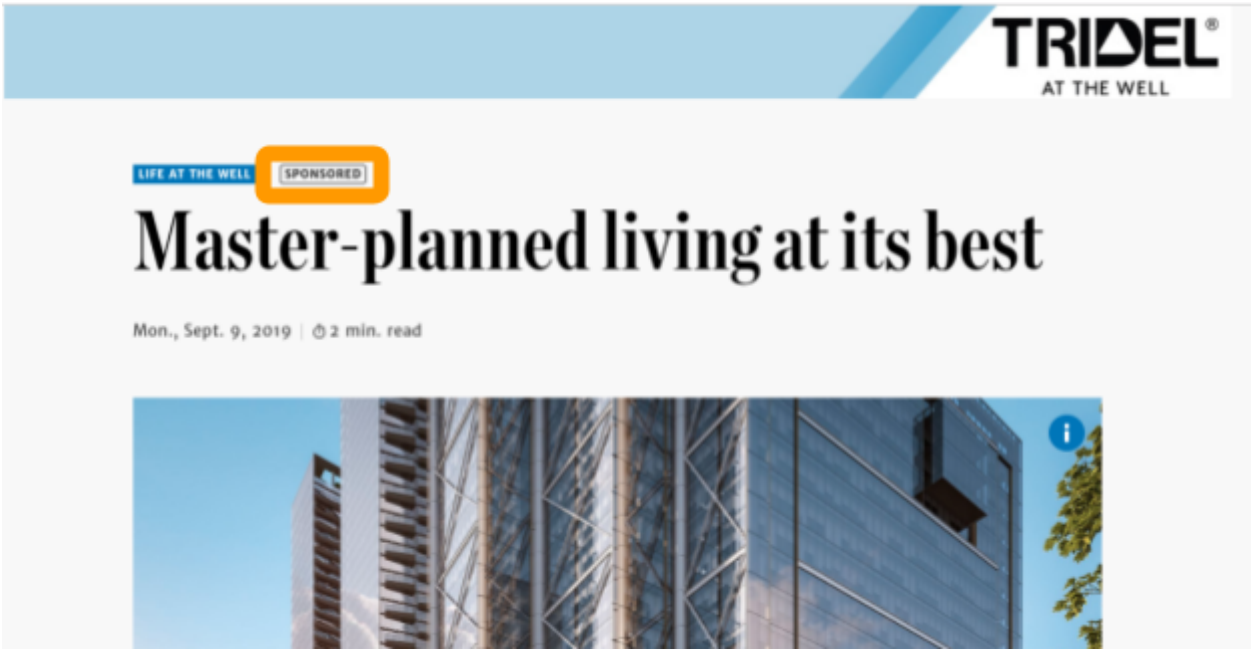
Type of Work

Este es un indicador orientativo de la TTP cuya finalidad es ser una etiqueta que permite distinguir distintos tipos de trabajo (como análisis, opinión, sátira, etc.) en distintos tipos de formato (como imagen, video, texto, entre otros). Esta información se debe utilizar en cada uno de los artículos publicados y su propósito es el de generar claridad sobre el tipo de contenido que se publica, Masera et al. (2018).

El artículo oficial sobre este indicador divide las etiquetas en tres grandes grupos, según Mesera et al. (2018): el primero es “noticias y opinión”, que incluye contenido de noticias (hechos verídicos y observables), artículos noticiosos (contenido noticioso como puede ser una página que lista las acciones en el mercado de valores), análisis (interpretación de las noticias basado en evidencia) y opinión (ideas y conclusiones basadas en la información disponible). El segundo grupo es “contenido no noticioso” y agrupa el contenido publicitario (contenido que produce y facilita un tercero que paga para que sea publicado), contenido patrocinado (contenido que produce la editorial en nombre de un patrocinador), y contenido con apoyo (contenidos que financia un tercero que quiere generar mayor interés sobre un tema específico o asociar una marca con ciertos temas). Finalmente están los “tipos opcionales” que incluye al contenido explicativo (contenidos enfocados en entregar información detallada sobre un tema en particular), revisión de hechos (verificar si una afirmación es cierta o no), “ayúdanos a informar” (cuando solicitan al público general compartir sus propias historias o hechos), obituarios, reseñas, investigativo (un examen a profundidad con investigación extensiva de un tema), detrás de la historia (aclarar al público cómo una historia fue realizada), sátira (uso de exageración, humor o ironía en un contenido).

Figura 2

Artículo de ejemplo de The Star



Fuente: Imagen extraída del artículo

https://www.thestar.com/sponsored_sections/life-at-the-well/2019/09/09/master-planned-living-at-its-best.html

Normalmente el tipo de contenido es escrito en una parte visible del contenido, en donde lo usual es resaltar el tipo de contenido según los lineamientos de TTP. También existen datos estructurados que ayudan a los robots a distinguir el tipo de contenido. En este caso, para el artículo de thestar.com vemos cómo resalta el término "*Sponsored*" -patrocinado- y a su vez explícitamente se declara que es un artículo publicitario en sus datos estructurados.


```
"@context":"http://schema.org",
"@type":"AdvertiserContentArticle",
"description":"Lots of planning goes into creating a community that will attract buyers and renters. It's called master-planned living.",
"articleSection":"Life at The Well",
"datePublished":"2019-09-09",
"dateModified":"2019-09-09",
```

```
"mainEntityOfPage":"https://www.thestar.com/sponsored_sections/life-at-the-well/2019/09/09/master-planned-living-at-its-best.html",

"headline":"Master-planned living at its best",
```

Fuente: Código extraído de *https://www.thestar.com/sponsored_sections/life-at-the-well/2019/09/09/master-planned-living-at-its-best.html*

Nuevamente los datos estructurados de TTP dependen de las buenas intenciones de los medios de comunicación, la presencia o ausencia de esta información no garantiza ni facilita la detección de noticias falsas. El mejor uso de esta información es que los medios noticiosos pueden otorgar información explícita para que Google u otro buscador no exponga como noticias un contenido que no lo es, pero no es una herramienta destinada a la detección de contenido que esté diseñado para ser dañino y desinformativo.

Citations & References

Esta es una señal que permite identificar una investigación genuina y facilita el rastreo de fuentes originales para que los usuarios puedan acceder a más detalles asociados a un contenido. Las citaciones y referencias son menciones a fuentes externas y enlaces a las mismas; son clasificadas en tres tipos: contenidos internos y originales del editor, contenidos originales externos al editor y fuentes noticiosas secundarias (otros editores que exponen información adicional sobre un contenido o el intercambio de contenido entre dos editores). Todo esto lo describe Bale et al. (2018) en el documento oficial.

Este indicador no fue encontrado en datos estructurados; por tanto, desde la perspectiva de Google depende únicamente del análisis del enlazado interno y externo junto a la comprensión de entidades durante el proceso de "*parsing*" para determinar qué es una fuente y cómo se asocia con un contenido. Sullivan, D., & Illyes, G. (2019) hablan sobre esto y explican la interpretación de Google frente a las etiquetas de robots en los enlaces: Google reconoce los etiquetados de enlaces patrocinados, enlaces de contenido creado por usuarios y los tradicionales "*follow*" y "*nofollow*" (que indican si Googlebot debe seguir un enlace o no). Esta información está alineada con TTP y permite concluir que los enlaces a

fuentes informativas es uno de los indicadores principales para determinar la importancia de un contenido y su probabilidad de ser un contenido dañino.

Methods

Este indicador ayuda a determinar si una pieza de contenido tiene alta calidad y originalidad, Myschasuk et al. (2018). Los autores especifican que los atributos del indicador "*Methods*" incluyen aspectos como el lugar y fecha de los hechos, fuentes y metodología para el desarrollo de un contenido, identidad del autor y su pericia temática y finalmente la identidad de los verificadores de hechos junto a la política de verificación de datos del medio de comunicación.

No vemos que este indicador especifique información adicional a la que ya podemos encontrar en el indicador de "*Best Practices*" y "*Journalist Expertise*" lo que dificulta ver su valor agregado dentro del análisis de la información. Es cierto que puede ayudar a especificar algunos datos sobre un contenido concreto como la metodología, pero esta información puede ser incluida fácilmente en el indicador de "*Best Practices*". Nuevamente la información contenida en este indicador no garantiza que una información sea verídica, pero ayuda a terceros a ser críticos con un contenido en concreto. El objetivo de este indicador parece estar orientado al análisis manual de información.

En términos generales, este indicador evalúa elementos ya incluidos en otros, la diferencia principal está en que la información no solo es verificada dentro del sitio, también es analizada en relación a una pieza de contenido individual, es decir, lleva señales de análisis para ser utilizadas en un artículo en concreto.

Local Reporting Indicator

La función de este indicador es la de exponer señales para determinar si una historia tiene relevancia local. Se basa en solo cuatro señales que ya fueron mencionadas en otros indicadores: información sobre la prioridad de cubrimiento de información local (antes mencionado en "*Best Practices*"), experiencia local (información atribuida al autor), especificar si el video de la historia es material local (antes especificado en "Methods"), y por último, la localización de la historia, Mauschard, B., Ackermann, S., Porto, M., & Mungeam, F. (2017)

En este indicador están siendo evaluadas otras señales ya utilizadas antes, pero analizadas desde una perspectiva local, con el fin de evaluar un contenido en concreto.

Diverse Voices

Este indicador es el encargado de referir la conciencia e inclusión de la diversidad en cada aspecto de la producción de un contenido noticioso y sus señales se incluyen en otros indicadores como "*Best Practices*", Hsu et al. (2017). Los autores hablan de distintos aspectos ya mencionados en otros indicadores de TTP como la declaración de diversidad. El atributo único encontrado es el que habla de las prácticas específicas que demuestren el compromiso con la diversidad de una editorial o contenido.

Este indicador es uno de los más difíciles de computar para Google y requiere presuntamente un gran volumen de revisión manual, ya que las prácticas mencionadas en el documento no hacen referencia a información externa o datos estructurados que permitan conocer con exactitud el cumplimiento de una compañía editorial frente a este indicador de confianza. Incluso hasta la fecha el documento oficial solo hace mención de algunas sugerencias que permitan evaluar este indicador de confianza, pero no menciona una metodología existente y vigente para ello. Por otra parte, no encontramos que este indicador tenga un peso considerable a la hora de evaluar información falsa: su propósito principal es el de garantizar la presencia de otros puntos de vista, lo cual puede evitar la existencia de información incompleta, pero no demuestra si un contenido es verdadero o falso.

Este por tanto no es un indicador importante para procesar o considerar a nivel semántico, luce como un lineamiento general que un buen sitio de noticias debería de seguir, no como un factor técnico a implementar dentro de una web.

Actionable Feedback

El último de los ocho indicadores de confianza de la TTP es un grupo de señales que permiten determinar si una editorial está comprometida con la escucha y profundización de los comentarios y retroalimentaciones de sus lectores, Terhaar, J., Haymarket, M. P., & Koon, B. (2017). Este indicador según el documento oficial se concentra en conocer el proceso utilizado por el publicador para recibir la retroalimentación de sus usuarios así como los canales disponibles para ello (sean foros, formularios de contacto, redes sociales

del autor, encuestas, votaciones, etc.). Por lo que el enfoque de este indicador más allá de conocer la existencia de una política de comunicación, busca exponer el uso y los resultados de la misma.

Qué significa *The Trust Project* para Google

The Trust Project es una iniciativa que apoya Google con el fin de combatir la información falsa. Los aportes de TTP a los medios periodísticos son utilizados por Google tanto a nivel cualitativo como cuantitativo. Cheng (2017) explica que la información de los indicadores de confianza de TTP puede ser insertada en un formato de Schema facilitando el trabajo de Google para analizar la información de cada artículo de distintas fuentes confiables, esto lo ha sido evidenciado en la evaluación técnica de distintos sitios web y además de corroborado que técnicamente es factible extraer y utilizar esta información con dichos fines. Está claro que TTP es fundamental para Google en el proceso de comprender y catalogar noticias, pero la forma en la que actualmente el proyecto funciona no sirve como una garantía o método eficaz para detectar *fake news,* sino que es un método que permite clasificar de una forma más exacta la información que ya publican las fuentes solventes.

No evidenciamos ningún indicio de que TTP permita fortalecer la exposición de medios independientes y fuentes de información más pequeñas ya que TTP depende de la confianza de la fuente. Solo hay dos escenarios posibles: o Google interpreta de forma neutral los datos estructurados relacionados con los indicadores de confianza de TTP presentes en todos los portales noticiosos sin importar su relevancia o reconocimiento, o Google solo tiene en cuenta los datos estructurados de los sitios aceptados por TTP. En cualquiera de los dos escenarios existe un problema: en el primer escenario no existe una forma en la que se pueda verificar la exactitud de la información ya que cualquiera puede modificarla a su antojo y en el segundo escenario la capacidad de difundir información y lograr exposición en Google será de los medios más solventes, de los medios que TTP aprueba siendo esto aún un aspecto que no garantiza la veracidad de la información.

Hay otros escenarios en donde los datos estructurados son importantes para Google, por ejemplo, Anderson (2017) explica en su artículo “Building trust online by partnering with the International Fact Checking Network” que Google tiene una alianza activa con la IFCN

con el objetivo de aumentar el número de personas verificadas dedicadas a la verificación de hechos, expandir el proyecto a más regiones y otorgar herramientas de verificación de datos de manera gratuita. Esto demuestra la incapacidad actual de Google para determinar qué es cierto y qué no lo es, y que necesita alianzas con instituciones especializadas y profesionales en todo el mundo.

Considerando lo que es TTP y sus alcances técnicos, concluimos que su valor se encuentra realmente en la estandarización de información ya que la posibilidad de comprender todos los aspectos relacionados al buen periodismo no son algo que pueda verse de forma explícita con solo los artículos noticiosos, es necesario contar con un respaldo tecnológico para que los robots puedan extraer información fundamental como el autor, las políticas de actuación, las colaboraciones entre medios, etc. TTP es un método que Google utiliza para analizar y alimentar a sus algoritmos de forma constante con esta información. Sin embargo, en lo que respecta a la información falsa no encontramos mecanismos o tecnologías que faciliten su detección; es decir, TTP no otorga métodos directos diseñados para combatir la desinformación sino que ofrece métodos para rastrear y reconocer con mayor facilidad los medios de comunicación más solventes y a los periodistas más reconocidos, sin que pueda ayudar a reconocer de forma alguna que uno de estos está publicando una *Fake New*.

¿Cumplen los periódicos con la implementación de estas señales?

Está claro que *The Trust Project* es una herramienta importante para Google para la detección de noticias falsas ya que aunque este gigante tecnológico cuenta con la tecnología necesaria para extraer y procesar información -el llamado proceso de *Parsing*- la estandarización de dicha información -en formato JSON-LD- es un enorme facilitador.

Esto implica que el valor real de *The Trust Project* está determinado por dos aspectos: en primer lugar contar con instrucciones claras, terminología estándar y además mecanismos escalables para su uso en cualquier sitio web; este aspecto el proyecto lo cumple con creces y como ha mostrado esta investigación en definitiva están aunados todos los datos necesarios para evaluar la fiabilidad de un medio de comunicación. El segundo aspecto es

el uso generalizado de estas características ya que, aunque se cuente con un estándar en la industria que todos puedan implementar de forma libre, si nadie lo implementa no servirá de nada; es decir, el éxito de *The Trust Project* depende precisamente de su acogida en los medios de comunicación.

Para evaluar esto tomamos un grupo de 30 periódicos españoles -los que registran más lectores diarios según la EGM- y evaluamos si en sus sitios web cumplen con 4 aspectos importantes de *The Trust Project*: 1. Si incluye de forma visible el logo del proyecto; 2. Si incluye la información mínima necesaria para cumplir con la señal de "Best Practices"; 3. Si cuenta con la información detallada de los autores y periodistas y finalmente; 4. Si incluye una declaración transparente del tipo de contenido publicado.

Tabla 1

Cumplimiento de las señales de The Trust Project por periódico

	Logo de TTP	Best Practices	Journalist Expertise	Type of Work
https://elpais.com/	Incluído	Cumple	Cumple parcialmente	Cumple parcialmente
https://www.20minutos.es/	Incluído	Cumple	Cumple	Cumple parcialmente
https://www.elmundo.es/	Incluído	Cumple parcialmente	Cumple	Cumple parcialmente
https://www.lavanguardia.com/	No incluído	No cumple	Cumple parcialmente	Cumple parcialmente
https://www.elperiodico.com/	No incluído	Cumple	Cumple	Cumple
https://www.abc.es/	No incluído	No cumple	Cumple parcialmente	Cumple parcialmente
https://www.lavozdegalicia.es/	No incluído	No cumple	Cumple parcialmente	Cumple parcialmente
https://www.elcorreo.com/	No incluído	No cumple	Cumple parcialmente	Cumple parcialmente
https://www.lne.es/	No incluído	No cumple	No cumple	Cumple parcialmente

	Logo de TTP	Best Practices	Journalist Expertise	Type of Work
https://www.heraldo.es/	Incluído	Cumple	No cumple	Cumple
https://www.levante-emv.com/	No incluído	No cumple	No cumple	Cumple parcialmente
https://www.farodevigo.es/	No incluído	No cumple	No cumple	Cumple parcialmente
https://www.larazon.es/	No incluído	No cumple	No cumple	Cumple parcialmente
https://www.laverdad.es/	No incluído	No cumple	Cumple parcialmente	Cumple parcialmente
https://www.diariovasco.com/	No incluído	No cumple	Cumple parcialmente	Cumple parcialmente
https://www.ultimahora.es/	No incluído	No cumple	Cumple parcialmente	Cumple
https://www.elnortedecastilla.es/	No incluído	No cumple	Cumple parcialmente	Cumple parcialmente
https://www.eldia.es/	No incluído	No cumple	No cumple	Cumple parcialmente
https://www.ideal.es/	No incluído	No cumple	Cumple parcialmente	Cumple parcialmente
https://www.diariodenavarra.es/	No incluído	No cumple	No cumple	Cumple parcialmente
https://www.eldiariomontanes.es/	No incluído	No cumple	Cumple parcialmente	Cumple parcialmente
https://www.laprovincia.es/	No incluído	No cumple	No cumple	Cumple parcialmente
https://www.diariosur.es/	No incluído	No cumple	Cumple parcialmente	Cumple parcialmente
https://www.elpuntavui.cat/	No incluído	No cumple	No cumple	Cumple
https://www.elcomercio.es/	No incluído	No cumple	No cumple	Cumple parcialmente

	Logo de TTP	Best Practices	Journalist Expertise	Type of Work
https://www.canarias7.es/	No incluído	No cumple	No cumple	Cumple parcialmente
https://www.hoy.es/	No incluído	No cumple	Cumple parcialmente	Cumple parcialmente
https://www.lasprovincias.es/	No incluído	No cumple	Cumple parcialmente	Cumple parcialmente
https://www.diariodeleon.es/	No incluído	No cumple	No cumple	Cumple
https://www.diariodecadiz.es/	No incluído	No cumple	No cumple	Cumple parcialmente

Fuente: Elaboración propia

Como es observable en la tabla, al momento de realizar esta evaluación (septiembre de 2022) ninguno de los periódicos logró un cumplimiento pleno de las señales de *The Trust Project*. Incluso los periódicos con mayor cumplimiento como lo son El País, 20 Minutos, *El Periódico* y El Mundo no logran cubrir a plenitud todos los aspectos fundamentales necesarios. *El Periódico* en este caso sería el medio con el mejor cumplimiento a nivel general ya que al carecer únicamente del logo de *The Trust Project* es fácil entender que los aspectos técnicos y funcionales son plenamente usables por Google.

Si en España solo 4 de los 30 periódicos evaluados hacen un uso mínimamente viable de las características propias de *The Trust Project*, es imposible e inviable -por lo menos en el contexto de España- pensar en este proyecto y toda la información necesaria para su cumplimiento como algo mínimamente útil para Google.

Después de todo si solo 4 periódicos indican con precisión -y eso que con algunos cumplimientos parciales- a Google toda la información necesaria para analizar los artículos periodísticos disponibles y determinar la confianza y veracidad de la información pública le será en última instancia imposible hacer un uso de esta información para detener la distribución de *fake news* por varios motivos:

1. No existe información detallada y verificable sobre los autores de cada periódico por lo que Google no puede usar de una forma generalizada la detección de los autores, su reconocimiento y por tanto sus datos para determinar si una noticia es confiable o no.
2. No hay una declaración transparente del tipo de contenido, es decir, la gran mayoría de periódicos no dice con claridad si el contenido de sus sitios web es una noticia, un contenido patrocinado, una sátira, etc. Todos exceptuando casos puntuales declaran a cualquier contenido como "noticia" cuando claramente no lo es. Esto desvirtúa cualquier tipo de uso que Google pueda dar a la señal de "Type of Work".
3. La información necesaria para cumplir con las "Best Practices" de *The Trust Project* simplemente no es un aspecto generalizado entre los periódicos analizados, lo cual dificulta cualquier análisis manual que un analista pueda hacer en un futuro.

Esto hace cuanto menos cuestionable el uso y aplicación real que tiene *The Trust Project* como señal usable por Google para combatir la desinformación. Si son datos que no tienen uso, aplicación generalizada ni tampoco un incremento en su uso a través del tiempo entonces es simplemente una herramienta que no tiene valor -al menos en la actualidad-.

Lo anterior hace plausible pensar que Google depende muchísimo más de criterios propios para detectar noticias falsas, y lamentablemente dichos criterios no cuentan con una declaración pública concreta.

Google Jigsaw: desarrollo activo frente a las *fake news*

Otro frente activo contra la desinformación es la división Jigsaw[5] de Google, un equipo que se define a sí mismo de una forma concreta:

> *"We look for high-impact interventions, where focusing on helping a specific group of people—journalists, civil society, or activists, for example—makes the internet and society stronger and safer for everyone. Our focus areas address some of the most complex challenges facing open societies." [Nosotros buscamos intervenciones de alto impacto, donde enfocarse en ayudar a un grupo específico de personas -periodistas, sociedad civil o activistas, por ejemplo- hace que el Internet*

[5] https://jigsaw.google.com/

*y la sociedad sean más fuertes y seguras para todos. Nuestras áreas de enfoque abordan algunos de los retos más complejos que retan a las sociedades abiertas] (*Google (2022), *parr. 2*)

Jigsaw aparenta ser un proyecto en el que Google busca lograr una posición más activa frente a actores malintencionados en Internet, lo cual incluye a los autores de las *fake news*. Si bien en su descripción no se explica de forma exacta qué tipo de soluciones o actividades Jigsaw como incubadora tecnológica de Google puede lograr, es posible comprender su lineamiento general a través de los 4 experimentos disponibles en su web[6]: *Assembler*, *Deepfake Dataset*, *StyleGAN Detector* y *Disinformation Data Visualizer*.

Assembler y StyleGAN Detector

Se trata de un proyecto cerrado centrado en ayudar a verificadores de información y periodistas a detectar contenido manipulado, Jigsaw. (2022). Probablemente Assembler sea de los primeros proyectos aplicados en un campo de uso real que usa parte de la teoría de detección de *fake news* mencionado en el marco teórico; Precisamente Assembler menciona en su web la colaboración con distintas universidades e instituciones a través del uso de distintos algoritmos y técnicas de detección de *fake news*[7].

Hao (2020) menciona que Assembler como solución es un paso en la dirección correcta, sin embargo no es una solución definitiva en parte porque la herramienta no tiene la capacidad de detectar todas las técnicas de manipulación existentes y además requiere de una actualización constante para poder lograr un nivel de eficiencia aceptable. Davey, A. (2020) expone que herramientas como Assembler son importantes precisamente por la presión actual para la diferenciación entre imágenes reales y alteradas.

Sin embargo, Assembler hoy en día es un proyecto cerrado y es válido asumir que los aprendizajes del proyecto explican por qué no se continuó con su desarrollo:

Implícitamente se entiende que Assembler como proyecto, pese a cualquier logro interesante que pudiese tener, no fue más que un experimento temporal con un impacto que no fue más allá de algún artículo de prensa -por algo es un proyecto cerrado y sin avances-.

[6] https://jigsaw.google.com/issues/
[7] https://projectassembler.org/collaborators/

Al momento de escribir esta tesis no fueron encontrados registros de logros, avances o victorias reseñables contra las *fake news*, y esto es algo comprensible desde los aprendizajes publicados en su web oficial.

> *The manipulations journalists deal with are not always those in training datasets*
>
> *The Challenge:*
>
> *To develop machine learning models that are capable of detecting certain types of detection, you need to train the model using examples of images with that type of manipulation. During Alpha testing, we found that fact-checkers and journalists are often tasked with debunking images that are underrepresented in our training sets and therefore the detectors aren't always able to accurately identify manipulations in these types of images. Some of the tricky cases we've observed include images that are screenshots of other screenshots and images that have been severely downsampled (taking a high-definition, large image and making it small) or reformatted (for example, changing the image format from JPEG to PNG).*
>
> *The Approach:*
>
> *Identifying these gaps in the training set has allowed us to focus on sourcing example images that can be used to train existing models to be able to accurately detect these cases, as well as source additional detectors to cover these gaps. We hope that in doing so we'll be able to help improve these detectors, as well as contribute back to the broader industry.*
>
> *[Las manipulaciones con las que se enfrentan los periodistas no siempre son las de los conjuntos de datos de entrenamiento.*
>
> *El reto:*
>
> *Para desarrollar modelos de aprendizaje automático que sean capaces de detectar ciertos tipos de detección, debe entrenar el modelo utilizando ejemplos de imágenes con ese tipo de manipulación. Durante las pruebas alfa, descubrimos que los verificadores de hechos y los periodistas a menudo tienen la tarea de desacreditar las imágenes que están subrepresentadas en nuestros conjuntos de entrenamiento y, por lo tanto, los detectores no siempre pueden identificar con precisión las*

manipulaciones en este tipo de imágenes. Algunos de los casos complicados que hemos observado incluyen imágenes que son capturas de pantalla de otras capturas de pantalla e imágenes que han sido severamente reducidas (tomando una imagen grande de alta definición y haciéndola pequeña) o reformateadas (por ejemplo, cambiando el formato de imagen de JPEG a PNG).

El enfoque:

Identificar estas brechas en el conjunto de entrenamiento nos ha permitido centrarnos en obtener imágenes de ejemplo que se pueden usar para entrenar modelos existentes para poder detectar con precisión estos casos, así como detectar detectores adicionales para cubrir estas brechas. Esperamos que al hacerlo podamos ayudar a mejorar estos detectores, así como contribuir a la industria en general.] Jigsaw (2022).

Este aprendizaje permite entrever uno de los aspectos fundamentales planteados en el marco teórico: la escalada tecnológica. Claramente el hecho de que exista un detector de imágenes falsas no garantiza que los autores de las mismas no encuentren nuevos métodos para evitar precisamente su detección. Así mismo la información es algo que está vivo, cambia y está en constante actualización, un sistema de detección de imágenes falsas aún con tecnologías como el *Machine Learning* necesita de constante aprendizaje para adaptarse a nuevos contenidos.

Este tipo de exigencias hacen que sea necesario un mantenimiento constante de la tecnología que no es fácil de continuar en el tiempo, dificultando su sostenibilidad. Esto puede ser precisamente una hipótesis de por qué Assembler está cerrado en este año.

The small, low-resolution images that journalists often deal with pose unique challenges to the detection technology

The Challenge:

Many fact-checkers deal with low-resolution and small images, coming from social media and instant messengers. This poses unique challenges to the detection technology, which is generally more accurate when images are in their original format and quality.

The Approach:

We are integrating an image auto-upgrading process, powered by TinEye, a popular reverse image search provider, which takes original images and finds larger and/or better quality versions of them in an effort to ensure the best image possible is analyzed by the detectors.

[Las imágenes pequeñas y de baja resolución con las que suelen lidiar los periodistas plantean desafíos únicos para la tecnología de detección.

El reto:

Muchos verificadores de datos se ocupan de imágenes pequeñas y de baja resolución que provienen de las redes sociales y los servicios de mensajería instantánea. Esto plantea desafíos únicos para la tecnología de detección, que generalmente es más precisa cuando las imágenes están en su formato y calidad originales.

El enfoque:

Estamos integrando un proceso de actualización automática de imágenes, impulsado por TinEye, un popular proveedor de búsqueda inversa de imágenes, que toma imágenes originales y encuentra versiones más grandes y/o de mejor calidad en un esfuerzo por garantizar que los detectores analicen la mejor imagen posible.] Jigsaw (2022).

Un aspecto a considerar es que la mayoría de los modelos existentes para la detección de *fake news* ignoran precisamente la resolución de la imagen, o en otras palabras: los modelos de aprendizaje se basan en bases de datos e información existente que se encuentra en condiciones óptimas. Nuevamente, dado que la información es transformada contínuamente en Internet es común encontrar distintas resoluciones y variantes de una misma imágen.

Journalists need to clearly understand model outputs and performance to make decisions

We heard from our Alpha testers that they need to better understand the relative strengths and weaknesses of each detector. They want to know: which type of

> *manipulation is each detector good for and which manipulation types does this detector not help with? As a result, we are reworking our user interface to provide clearer explanations on individual detector performance and different detector results.*
>
> *[Los periodistas deben comprender claramente los resultados y el rendimiento del modelo para tomar decisiones*
>
> *Escuchamos de nuestros evaluadores Alpha que necesitan comprender mejor las fortalezas y debilidades relativas de cada detector. Quieren saber: ¿para qué tipo de manipulación es bueno cada detector y con qué tipos de manipulación no ayuda este detector? Como resultado, estamos reelaborando nuestra interfaz de usuario para brindar explicaciones más claras sobre el rendimiento de detectores individuales y los diferentes resultados de los detectores.] Jigsaw (2022)*

Según deja entrever este aprendizaje, las personas encargadas de usar a Assembler no lograron comprender bajo total claridad su uso, dificultando precisamente la efectividad de la herramienta. Debe considerarse que las técnicas existentes para detectar imágenes falsas no se basan en un único método o algoritmo, dado que los modelos de modificación y alteración de imágenes pueden tener una significativa diversidad metodológica es imprescindible hacer que la herramienta cuente con cierta flexibilidad -o bien crear múltiples herramientas en paralelo-. Si los operadores de la herramienta no conocen su funcionamiento técnico ni comprenden la mejor forma de usarlo entonces no será posible obtener un resultado realmente útil.

Assembler es un proyecto que demuestra en primer lugar la necesidad que tiene Google de tomar un rol activo en la detección de *fake news* ya que no es posible basarse únicamente en la información presente en periódicos confiables o en proyectos como *The Trust Project* para mitigar el efecto de la desinformación en la población. Así mismo este proyecto de forma implícita muestra que no consiguió los resultados esperados. Detectar *fake news* no es una tarea sencilla ni de una única solución.

Deepfake Dataset

En el año 2019 Google anunció en su blog oficial de Inteligencia Artificial la publicación de un conjunto de datos que contiene un número sustancial de Deepfakes creadas por Google, Dufour & Gully (2019).

Este conjunto de datos es de acceso público y puede ser descargado desde el repositorio oficial en GitHub[8], esto implica que el trabajo hecho por Google es más "filantrópico" que "comercial", es decir, el objetivo de Google con este proyecto no es el de mejorar las características intrínsecas de YouTube, Google Search o cualquier otro servicio de Alphabet para detectar contenido desinformativo, sino más bien ayudar a la academia a encontrar más y mejores métodos para ello. Aunque esto no es algo dicho de forma explícita por Google sí podemos leer en el artículo de Dufour & Gully (2019) que precisamente el uso de este conjunto de datos estuvo principalmente en la Universidad Federico II de Nápoles y en la Universidad Técnica de Munich.

A diferencia de Assembler, sí existen algunos documentos que evidencian el uso del *Deepfake Dataset*; por ejemplo, el FaceForensics Benchmark[9], aun así estos usos no terminan de ser ejemplificaciones meramente académicas, o en otras palabras: escenarios meramente teóricos.

Es de hecho llamativo que el *Deepfake Dataset* se encuentre con más de 2 años de inactividad al momento de escribir esta tesis doctoral (agosto de 2022). Este hecho solo demuestra el escaso uso práctico que este conjunto de datos tiene y que de hecho su información tiene una aplicación meramente académica y teórica. Sin duda un aporte valioso para las ciencias pero sigue sin ser una solución contundente usable por Google para la detección de contenido desinformativo.

Disinformation Data Visualizer

Este proyecto no es una solución tecnológica -lo que hace extraña su presencia en Jigsaw-, Jigsaw (2022) describe el *Disinformation Data Visualizer* de la siguiente manera: "This project visualizes the Atlantic Council's Digital Forensic Research Lab research on

[8] https://github.com/ondyari/FaceForensics/
[9] https://kaldir.vc.in.tum.de/faceforensics_benchmark/index.php

coordinated disinformation campaigns." [Este proyecto visualiza la investigación del Laboratorio de Investigación Forense Digital del Atlantic Council sobre campañas coordinadas de desinformación.]. Evidentemente esto no es más que un proyecto dedicado a visualizar la información de un tercero en la que Google no participó de ninguna manera.

En su página oficial[10] Jigsaw (2022) aclara de forma literal que "This project visualizes coordinated disinformation campaigns identified by the Atlantic Council's DFRLab. Alphabet does not endorse these research findings or their characterization of disinformation campaigns." [Este proyecto visualiza campañas de desinformación coordinadas identificadas por el DFRLab del Atlantic Council. Alphabet no respalda estos hallazgos de investigación ni su caracterización de campañas de desinformación.].

Este proyecto como tal deja muy claro que el rol de Jigsaw está limitado explícitamente a la visualización de datos y no a la detección de la información falsa presente en este proyecto, es más, Google o Jigsaw tampoco afirman o desmienten que la información de dicho proyecto sea una *Fake New* como tal. Claramente Google o no tiene la capacidad de detectar una noticia falsa o de forma arbitraria no quiere entrar en dicha tarea de clasificar un contenido.

En una cita más extensa es posible comprender en mayor detalle las intenciones de Google tras este proyecto:

> *This project visualizes the Atlantic Council's DFRLab research on coordinated disinformation campaigns. The campaigns included reflect DFRLab's own analysis and perspective, using their own tools and datasets or those obtained from their partners. Google does not endorse these research findings or their characterization of disinformation campaigns. Google combats coordinated disinformation campaigns across its products and its enforcement actions are driven by its own independent investigations, which include consideration of intelligence and data that is not available to external researchers. The Visualizer is intended to help newcomers better understand disinformation campaigns. DFRLab's research predominantly relies on open source, English language press reporting of coordinated disinformation campaigns that may appear to target the West. [Este*

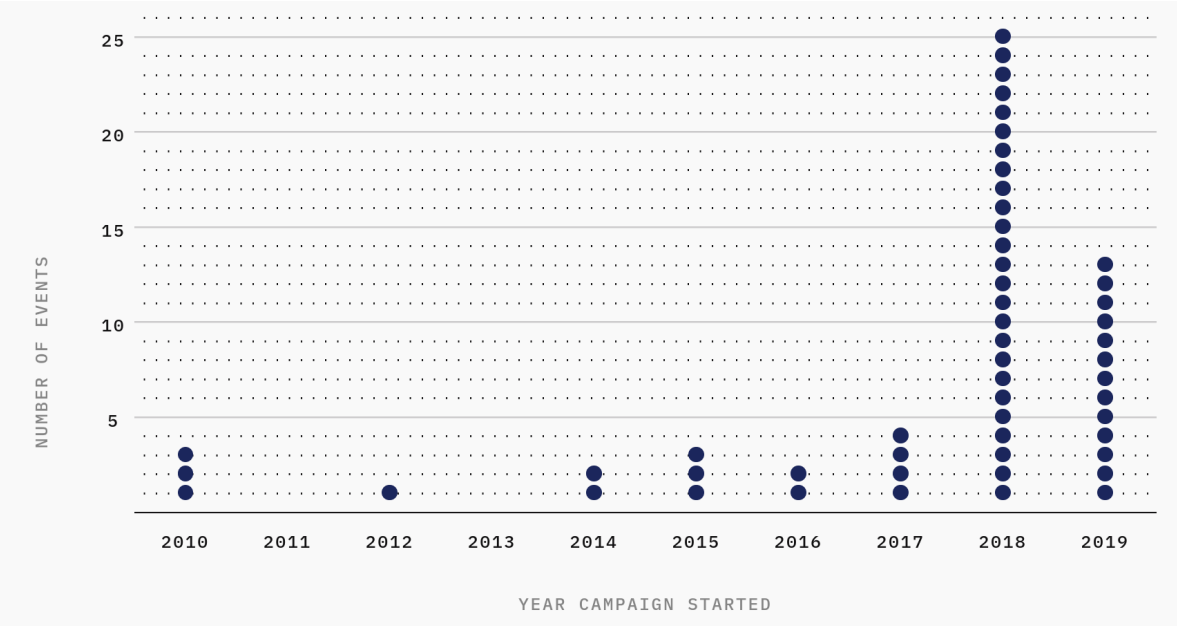
[10] https://jigsaw.google.com/the-current/disinformation/dataviz/

proyecto visualiza la investigación DFRLab del Atlantic Council sobre campañas de desinformación coordinadas. Las campañas incluidas reflejan el análisis y la perspectiva propios de DFRLab, utilizando sus propias herramientas y conjuntos de datos o los obtenidos de sus socios. Google no respalda estos hallazgos de investigación ni su caracterización de campañas de desinformación. Google combate las campañas de desinformación coordinadas en todos sus productos y sus acciones de cumplimiento están impulsadas por sus propias investigaciones independientes, que incluyen la consideración de inteligencia y datos que no están disponibles para investigadores externos. El visualizador está destinado a ayudar a los recién llegados a comprender mejor las campañas de desinformación. La investigación de DFRLab se basa predominantemente en informes de prensa en inglés de fuente abierta sobre campañas de desinformación coordinadas que pueden parecer apuntar a Occidente.] Jigsaw (2022).

Es interesante que en la propia descripción del proyecto Google especifique que en sus productos sí que aplica técnicas de detección de noticias falsas, no obstante ninguno de los productos tecnológicos de Jigsaw figuran en dichas técnicas.

Figura 3

Línea de tiempo publicada en Disinformation Data Visualizer



Fuente: Jigsaw (2022)

Llama la atención que este proyecto, al igual que los anteriormente mencionados, denote un nivel de inactividad prolongado ya que las últimas noticias falsas detectadas datan del año 2019, es decir, este producto se encuentra presuntamente inactivo y no existe comunicación pública de Jigsaw o Google sobre futuras actualizaciones del mismo. En todo caso esta herramienta en sí mismo no es útil para la detección de *fake news* en un entorno digital, solo es útil para su estudio y análisis.

Factores E-A-T

Como se ha visto anteriormente, Google participa activamente en distintos proyectos potencialmente útiles para la detección de *fake news,* aunque su uso tangible en entornos prácticos y no meramente teóricos son difíciles de evidenciar, incluso la evidencia escrita dice implícitamente que todo ha tenido un uso limitado al entorno académico o político -no hay un caso registrado en el que estas iniciativas hayan ayudado a combatir a las *fake news*-. Como último recurso investigativo quedan las actualizaciones algorítmicas

registradas que permitan de forma indirecta entender qué tipo de acciones toma Google para precisamente detectar una *Fake New.*

En este caso la información pública es limitada y el único factor explícito detectable es el E-A-T, un conjunto de señales que traducen como "Expertis, Autoridad y Confianza -*trustworthiness*-", Crowe (2021).

La teoría actual sobre las señales E-A-T proviene del documento guía general de los revisores de contenido de Google[11] y Google (2022) lo explica de la siguiente manera:

> *The amount of expertise, authoritativeness, and trustworthiness (E-A-T) is very important. Please consider:*
>
> ● *The expertise of the creator of the MC -Contenido principal-.*
>
> ● *The authoritativeness of the creator of the MC, the MC itself, and the website.*
>
> ● *The trustworthiness of the creator of the MC, the MC itself, and the website.*
>
> *[La cantidad de experiencia, autoridad y confiabilidad (E-A-T) es muy importante. Por favor considera:*
>
> ● *La experiencia del creador del MC -Contenido principal-.*
>
> ● *La autoridad del creador del MC -Contenido principal-, el propio MC -Contenido principal- y el sitio web.*
>
> ● *La confiabilidad del creador del MC, el propio MC -Contenido principal- y el sitio web.]*

Según esta señalización de Google, la reputación del autor es clave para determinar la confianza que se puede depositar sobre un contenido, aspecto que encaja con la información que *The Trust Project* pide a los periódicos añadir de forma proactiva en sus contenidos, después de todo entre más fácil sea para Google detectar a un autor le será más sencillo saber si se trata de una persona confiable.

> *High E-A-T news articles should be produced with journalistic professionalism –they should contain factually accurate content presented in a way that helps users*

[11] https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf

> *achieve a better understanding of events. High E-A-T news sources typically have published established editorial policies and robust review processes [Los artículos de noticias de alta E-A-T deben producirse con profesionalismo periodístico: deben contener contenido preciso y presentado de una manera que ayude a los usuarios a lograr una mejor comprensión de los eventos. Las fuentes de noticias de alta E-A-T generalmente han publicado políticas editoriales establecidas y procesos de revisión sólidos.]. Google (2022)*

Para el caso de las noticias y el contenido periodístico Google hace especial hincapié en el profesionalismo bajo el cual un contenido ha sido creado, nuevamente hace referencias indirectas a contenidos que deberían ser publicados si un portal de noticias sigue los lineamientos provistos por The Trust Project. De alguna forma lo que Google hace ver aquí es que el análisis de los factores de TTP son precisamente revisados por un ser humano para determinar el valor E-A-T de un sitio web, aspecto que en última instancia debería ser utilizado para resolver el posicionamiento en el motor de búsqueda de Google.

Conclusiones

El panorama de Google frente a las noticias falsas se puede dividir en dos puntos clave: 1) aplicación tecnológica de estándares informativos y 2) la creación de alianzas y programas periodísticos. Esto quiere decir que Google comprende que hay dos frentes diferentes que deben interactuar entre sí con el fin de combatir las noticias falsas: tecnológicos e institucionales.

El frente tecnológico necesita de una estandarización de datos sólida que permita alimentar con información detallada y a bajo coste computacional a las herramientas de Google, por ello sus esfuerzos incluyen la incentivación del uso de estándares informativos como *Schema*, de forma que Google pueda entender el contenido online y así fortalecer su comprensión de las entidades.

En el frente institucional Google necesita que profesionales de todo el mundo utilicen sus herramientas tecnológicas y adopten sus estándares para facilitar el trabajo de detección de noticias falsas, pero también necesita alianzas estratégicas que le permitan abarcar la

verificación de hechos a nivel global y local. Esto también incluye prácticas éticas de periodismo, inclusión y transparencia.

Google no tiene herramientas tecnológicas exactas que detecten las "*fake news*" y *The Trust Project* como principal aliado no tiene la madurez técnica de facilitar estándares útiles para todas las voces que existen y que son diferentes a los medios de comunicación más solventes. La exposición en Google Search y Google News es filtrada principalmente por la reputación y confianza de la fuente para evitar la presencia de información dañina para el usuario, lo que no es una garantía de que la información expuesta sea verídica.

Es cierto que Google ha invertido de distintas maneras en metodologías y herramientas útiles para la detección de las *fake news* aunque la mayoría de éstas han tenido un efecto mediático en el sentido que atraen noticias favorables para Google y no un efecto práctico en la necesidad actual que tienen los medios tecnológicos para detectar información falsa a gran escala. Es factible pensar que Google continuará invirtiendo en nuevas iniciativas aunque la tasa de éxito sea baja.

Está claro que Google depende principalmente del proceso de *parsing* para extraer el contenido de un sitio web y es un hecho que utiliza distintos métodos computacionales complejos para determinar qué tipo de información tiene una URL en concreto -especialmente bajo técnicas de NLP o *Natural Lenguaje Processing*-. Aún así aunque estas técnicas sean muy complejas, Google depende de una estandarización clara y escalable para comprender la información presente de forma precisa y económica -ya que procesar información tiene un costo energético considerable-. Eso lleva a la conclusión de que *The Trust Project* es la apuesta más lógica de Google para lograr comprender qué contenido es confiable y por qué.

Finalmente, a partir de los datos analizados también se concluye que *The Trust Project* como herramienta fundamental para facilitar la detección y análisis de los contenidos periodísticos no tiene un uso generalizado en España y por tanto su uso práctico en este país no debería generar ningún valor añadido para Google. Asumiendo que solo Estados Unidos es el país en el que los periódicos han hecho un uso adecuado de los estándares de este proyecto entonces lo mejor que puede decirse es que *The Trust Project* tiene un uso

localizado, el cual no tiene aplicación en ninguna otra región. Estará bien contar con futuras investigaciones que evalúen el uso e impacto real de *The Trust Project.*

Otro aspecto fundamental es la poca adopción generalizada -en España- de los estándares de *The Trust Project*, y esto es crítico ya que aunque Google promueva el uso de un estándar de datos perfecto si nadie lo usa entonces Google -ni ninguna otra compañía- podrá hacer uso de los datos. Esto es importante ya que los estándares como *The Trust Project* no son útiles porque sean precisos o detallados, son útiles porque muchas instituciones los usan y hace sentido invertir en su interpretación. Este no parece ser el caso y hace que sea importante cuestionarse el futuro de estas metodologías al momento de detectar una *Fake New.*

Pese a lo anterior es notable la dificultad que Google tiene para crear herramientas definitivas para la detección de *fake news* y de hecho su principal incubadora tecnológica para hacer frente a este fenómeno -Jigsaw- cuenta con distintos proyectos inactivos cuyos resultados son muy limitados.

Todo apunta a que Google se encuentra en un proceso de mejora constante en donde cada vez depende más de la Inteligencia Artificial y el *Machine Learning* para detectar y comprender cada tipo de contenido, sin embargo su dependencia actual en lo que corresponde a la estandarización de contenidos y formatos -JSON y Schema- dificultan notoriamente el desempeño de sus sistemas de detección.

A priori parece que Google depende más de la revisión humana y de factores como los llamados E-A-T para determinar si una fuente de contenidos es confiable, pero de ninguna manera para saber si una información es cierta o no.

Sencillamente Google depende en gran medida de procesos humanos y la construcción de métodos maduros que usen señales concretas para detectar una *Fake New* está lejos de concretarse. Para investigaciones futuras será importante retomar este tema concreto y evaluar la adopción de los estándares existentes -TTP- o de nuevos estándares por parte de Google.

Esto no descarta que en el futuro estos estándares sean la herramienta más fiable para grandes tecnológicas -aparte de Google- para detectar qué contenido es falso, o al menos

poder calificar cada una de las fuentes de información según su confianza y reputación. Hay que destacar la utilidad de las herramientas y estándares propuestas pese a su poco uso y evolución reciente. En parte esto ocurre porque la implementación de todas estas herramientas no depende únicamente de Google.

GOOGLE Y EL FILTRO DE LAS *FAKE NEWS*: EL CASO DE LAS ELECCIONES PRESIDENCIALES DE LOS ESTADOS UNIDOS EN 2016

Introducción

Las elecciones presidenciales del año 2016 en Estados Unidos son un punto de inflexión en la historia de las *fake news* (Alfonso, Galera y Calvo, 2019) en donde podemos evidenciar el poder de la información falsa, no solo por su capacidad de manipular, sino por su enorme distribución a través de medios digitales como Google o Facebook. Después de todo estos son medios propicios para su distribución, Allcott & Gentzkow (2017). Google y Facebook son un duopolio con una audiencia tan grande que los convierte en puntos llamativos para quienes producen *fake news*, ya que estas necesitan distribución con una gran audiencia para cumplir su cometido lo que otorga a las grandes tecnológicas un gran poder y una gran responsabilidad: de repente el mundo entero depende de empresas privadas como Google y Facebook -dos de las empresas con una posición dominante en el mercado, Haucap & Heimeshoff (2014)- para evitar la desinformación y la manipulación.

Google es una de estas grandes tecnológicas cuyo buscador alcanza más de 83 mil millones de visitas al mes (SimilarWeb, 2020) siendo esta empresa la que controla qué aparece en su buscador y cómo aparece, por tanto es una de las compañías privadas que tiene el poder de moderar y limitar la distribución de las *fake news*. En lo que respecta a la moderación de contenidos digitales en este momento se trata de una decisión que implica varias definiciones como "qué es verdad", "qué es contenido malicioso" y "qué es considerado manipulación"; dichas definiciones no las toma un gobierno o una población democráticamente, es una decisión unilateral por parte de compañías privadas ya que estas definiciones hacen parte de las políticas de términos y condiciones privadas -de Google- sin que exista una ley o punto en común para definir estos aspectos; es decir, Google modera el contenido desinformativo según sus propios criterios de lo que es desinformativo. No solo eso, también la tecnología de sus algoritmos es opaca, nadie -externo a Google- sabe cómo funcionan exactamente ni nadie puede decir que sus filtros cumplan con las expectativas

éticas de una población en concreto. Estamos hablando de una compañía norteamericana cuyas decisiones tienen implicaciones a nivel global económica, política y socialmente.

Con este contexto en mente en este texto estudiamos el impacto de Google sobre sitios web dedicados a la distribución de *fake news* tras los resultados electorales del año 2016 en los Estados Unidos. Dado que no es posible especificar con exactitud qué cambios algorítmicos existieron -sino que solo es posible referenciar "posibles cambios" a partir de datos observables- nos enfocaremos en los resultados en el posicionamiento orgánico y en las decisiones que Google toma tras dichas elecciones. Es decir: nos dedicamos a cuantificar, analizar y describir qué sucedió a partir del año 2016. Como unidad de análisis tendremos a la lista conocida de dominios que Google de forma manual excluyó de su corpus de noticias en Google Now[12] con el objetivo de encontrar patrones que nos ayuden a entender si tras las elecciones de 2016 Google ejecutó cambios orientados a combatir la desinformación, o bien, un tipo específico de contenido.

El objetivo de este análisis es entender qué cambios, prácticas y tecnologías aplica Google para lograr localizar y filtrar el contenido desinformativo en sus resultados de búsqueda, así mismo cuantificar a través de herramientas profesionales el impacto de estos cambios.

Metodología

Este artículo es un análisis estadístico de un grupo específico de sitios web -a los que también nos referiremos como "dominios"- incluídos en una lista privada creada por Google y que Project Veritas filtró al público[13]. Según varios reportes como el de Fruen, L. (2019) explican que esta lista fue utilizada para excluir sitios web con el fin de limitar su aparición en Google Now -según Wordstream (n.d.) Google Now es una aplicación hecha para personalizar la experiencia de los usuarios usando tarjetas, entre ellas las de "últimas noticias"-. Dentro del archivo original encontramos un total de 492 sitios.

Metodológicamente no nos interesa la opinión de Project Veritas o de cualquier otro autor u organización, solo nos concentramos en el hecho de que la lista es real y que además cuenta con comentarios que explican vagamente cuáles son los motivos por los cuales excluyeron

[12] https://www.thesun.co.uk/news/9726590/google-blacklist-conservative-whistleblower/
[13] https://pv-uploads1.s3.amazonaws.com/uploads/2019/08/news-black-list-site-for-google-now.txt

cada uno de los dominios. Nos llama la atención que en el archivo de texto original existan comentarios referentes a sitios que los usuarios "bloquean frecuentemente" y sitios reportados por distribuir información falsa -hoaxes-. Esto hace de la lista un objeto de estudio interesante para comprender el impacto de Google sobre el posicionamiento de contenido desinformativo -específicamente el contenido que Google considera como desinformativo-.

Este grupo de dominios es basto y temáticamente diverso, algunos de ellos son tiendas (otakumode.com), otros ofrecen información profesional (blackhatworld.com), entre otros. Lo que sí es cierto es que la mayor parte de ellos son portales noticiosos -información que demuestra una parcialidad en contra de los sitios web conservadores según Project Veritas, Lakshmanan (2019)- por ejemplo, bizpacreview.com. Considerando esto para entender el objeto de estudio es necesario tener en cuenta:

1. La muestra que compone el objeto de estudio -los dominios bloqueados por Google- pese a estar compuesta en su mayoría por sitios noticiosos, también incluye sitios de otras categorías como foros, comercio electrónico, entretenimiento, etc. No excluimos ningún sitio con el fin de no alterar los resultados debido a que el objetivo es estudiar el posicionamiento histórico de todos los sitios que Google ha bloqueado a través de esta lista. En su lugar, a través del análisis de datos realizaremos distintos filtros para entender cómo determinadas categorías se comportan en comparación a otras.
2. Dada la amplitud de la muestra y su variedad de información, decidimos acceder a la base de datos de zvelo.com -una compañía que ofrece un servicio de clasificación de URL's- para encontrar la categoría exacta de cada sitio web. Esto principalmente porque una parte considerable de los sitios se encuentra inactivo -motivo por el cual la información presente en Zvelo es de las pocas que puede dar una categorización precisa-. Por otra parte, los más de 20 años de experiencia con la que Zvelo cuenta categorizando todo tipo de webs y su seguimiento de estándares como los que propone el IAB -una reconocida asociación de comunicación, publicidad y marketing digital- otorgan la confianza suficiente como para utilizar su información de categorización en esta investigación. Esto además ayuda a garantizar un nivel de

neutralidad en este estudio ya que al existir una clara tendencia ideológica en varios de los sitios listados es importante contar con categorización neutral con el fin de evitar los juicios de valor personales. En otras palabras, Zvelo es un mecanismo neutral y profesional que nos otorga una posible clasificación de cada uno de los sitios que es útil para este estudio.

Se ha decidido no filtrar los 492 dominios debido a que hipotéticamente si todos los dominios fueron excluidos de la sección de noticias de Google Now -independientemente de su categorización o tipo de contenido- tendrán resultados históricos de posicionamiento similares en caso de que Google no tenga actualizaciones dirigidas a los sitios de noticias, por el contrario de encontrar resultados históricos diferentes entre categorías podremos tener una idea de qué cambios ha hecho Google y por qué.

Es importante considerar que aunque sean sitios de categorías diferentes podrían tener factores en común ya sea en su contenido o en la forma en la que Google los posiciona en su motor de búsqueda. Adicionalmente, todos están incluidos en el mismo filtro de Google. Por tanto, excluirlos basándonos en su contenido podría viciar los datos del análisis. En todo caso en la creación de algunos gráficos algunas categorías o sitios podrían ser filtrados para apreciar información más detallada y dichos filtros -de ser aplicados- serán detallados en cada uno de los gráficos.

El motivo por el cual este análisis se concentra en esta lista de sitios es debido a que es la única lista con participación mayoritaria de sitios web noticiosos en donde existe la posibilidad de notar una acción directa de Google en su posicionamiento y en la que además contamos con una verificación explícita de una intervención manual por parte de Google -los ha bloqueado manualmente para no aparecer en Google Now-. En contraste a esto un análisis exhaustivo de toda la red o incluso de todos los sitios web de noticias en Internet es simplemente una tarea que no es posible abarcar con los recursos disponibles. Dicho esto, contar con una lista específica de sitios a los cuales de forma explícita sabemos que Google ha rechazado en uno de sus servicios asociados a las noticias -sea por ser sitios de baja calidad, por ofrecer noticias falsas o no ofrecer contenido noticioso- nos permite entender el impacto a lo largo del tiempo de las actualizaciones de Google.

Tras extraer de Zvelo[14] la categorización de cada dominio siguiendo la Taxonomía del IAB[15] encontramos los siguientes grupos y tipos de categorización:

Tabla 1

Definición de tipo de contenido.

Contenido	Definición
Content Categories **Categorías de contenido**	Categorización principal de un sitio web. Un solo sitio puede pertenecer a una o a varias categorías. Ej: Red social, contenido educativo, noticias, etc. Zvelo dispone de más de 500 categorías.[16]
Iab-tier-1 **Grupo 1 IAB**	Categorización según las reglas taxonómicas del IAB. "Tier 1" corresponde a la categoría principal de la cual se desprenden distintas subcategorías. Por ejemplo: "libros y literatura" es una categoría "Tier 1".
Iab-tier-2 **Grupo 2 IAB**	Categorización según las reglas taxonómicas del IAB. "Tier 2" son categorías más precisas dentro de una categoría "Tier 1", por ejemplo "ficción" es una categoría "Tier 2" que pertenece a la categoría "libros y literatura".
Iab-content-rating **Clasificación de contenido IAB**	Categoriza un sitio web en relación a los grupos de edades apropiados para un contenido. Ej: "mayores de edad" o "todas las audiencias".

[14] https://tools.zvelo.com/
[15] https://www.iab.com/guidelines/taxonomy/
[16] https://zvelo.com/content-dataset/

Contenido	Definición
Iab-non-standard **IAB no estándar**	Este tipo de categorización del IAB reúne todos los sitios con categorías especiales[17] usualmente con connotaciones negativas como pornografía, lenguaje profano, contenido de odio, etc.
Iab-illegal-content **Contenido ilegal IAB**	Este tipo de categorización del IAB indica si un sitio presenta contenido ilegal como piratería, *Malware* o violación de copyright.
Brand-safe **Seguridad de marca**	Este tipo de categorización de Zvelo indica si es seguro para una marca anunciar en un sitio específico[18]. Solo se categoriza de dos formas: "sí" o "no".
Malicious **Malicioso**	Esta categorización de Zvelo determina si un sitio tiene contenido malicioso como virus informáticos, fraude publicitario, engaño, etc[19]. Solo se categoriza de dos formas: "sí" o "no".
Objectionable **Objetable**	Esta categorización indica si un contenido es apropiado para todas las audiencias o si presenta restricciones al ser *fake news*, contenido de odio u otro tipo[20].

Fuente: Elaboración propia con las definiciones existentes en los documentos oficiales del IAB y de Zvelo.

[17] https://web.archive.org/web/20171017061949/http://www.iab.net/media/file/NE-QA-Guidelines-Final-Release-0610.pdf
[18] https://zvelo.com/solutions/brand-safety-contextual-targeting/
[19] https://zvelo.com/solutions/malicious-detection
[20] https://zvelo.com/objectionable-dataset/

Para obtener datos cuantitativos de nuestra muestra decidimos tomar como fuente de información principal a Sistrix, una compañía alemana dedicada a almacenar y cuantificar el posicionamiento histórico de millones de sitios webs para millones de consultas de forma diaria. Consideramos esta web como una fuente fiable al ser una de las pocas que mide de forma consistente información detallada de Google desde el año 2008 según explican en su página "About Sistrix" Sistrix. (2020). Pocas herramientas entregan información tan consistente y con indicadores tan estables.

El indicador principal que estaremos utilizando es el denominado "índice de visibilidad" de Sistrix, el cual es una métrica calculada que Sistrix (2020) explica en su artículo "SISTRIX Visibility Index – Explanation, Background and Calculation" como un proceso de 3 partes: 1, búsqueda de las 100 primeras posiciones de todas las palabras clave existentes en su diccionario. 2, ponderar los resultados basándose en la probabilidad de clic según la posición y en el volumen de búsquedas que tiene cada una de las palabras clave. 3, sumar los resultados a cada uno de los dominios.

En otras palabras, el índice de visibilidad es un indicador que nos da una imagen general de cómo un sitio web -dominio- posiciona en Google, entre más alta su posición, más palabras clave posicionadas y más búsquedas tengan dichas palabras, más alto será su índice de visibilidad. Esto significa que un cambio abrupto de este indicador será un indicio claro de cómo un cambio en Google puede castigar o premiar ciertos sitios de nuestro objeto de estudio.

Además de Sistrix se ha incluído datos cuantitativos de Ahrefs[21], una compañía que también se dedica a medir de forma cuantitativa el posicionamiento orgánico en Google y los "backlinks" -por backlink podemos entender simplemente un enlace desde un sitio web a otro- de los dominios que estaremos evaluando dentro de este estudio. Por parte de Ahrefs a través de su panel de control descargamos información de los siguientes indicadores:

[21] https://ahrefs.com/big-data

Tabla 2

Definición de términos.

Indicador	Definición
Dominios de referencia	El número total de dominios únicos que enlazan a tu objetivo.
Páginas de referencia	Número total de páginas únicas que enlazan con un sitio web o URL de destino.
Clasificación del Dominio (Domain Rating - DR)	Muestra la solidez del perfil de backlinks de un sitio web objetivo en una escala logarítmica de 0 a 100, siendo este último el más fuerte.
Tráfico orgánico	Esta métrica estima la cantidad de tráfico de búsqueda orgánica que recibe el sitio web, dicha estimación se realiza en base a la cantidad de keywords posicionadas, su posición actual y su posibilidad de recibir un clic en el resultado.
Palabras clave orgánicas	El número total de palabras clave por las que se posicionan en los resultados de búsqueda orgánicos. Ahrefs busca hasta en los primeros 100 resultados a través de 605 millones de keywords

Fuente: Elaboración propia con las definiciones existentes en el control panel de Ahrefs [22] y en su artículo de definición de métricas[23].

[22] https://ahrefs.com/site-explorer/overview/v2/subdomains/live
[23] https://ahrefs.com/blog/seo-metrics/

Finalmente dicho todo lo anterior, ejecutamos un proceso detallado de descarga masiva de datos (limitándose a Estados Unidos siempre que fuese posible) a través de tres pasos: 1. Descargar todos los datos históricos semanales en los Estados Unidos contenidos en Sistrix a través de su API[24], lo que en palabras más simples quiere decir que solicitamos directamente a los servidores de Sistrix la información necesaria para este análisis y la organizamos en un archivo separado por comas -CSV- para poder analizarla. También descargamos manualmente 5 archivos CSV por cada uno de los dominios desde Ahrefs para incluir toda la información disponible desde este servicio. 2. Solicitar manualmente a Zvelo la información de cada uno de los sitios de forma que podamos contrastar mediante columnas no solo el índice de visibilidad de cada sitio, sino también el de distintas categorías. 3. Procesamos los datos de todos los archivos -más de 6 millones de filas solo en Ahrefs- y normalizamos sus formatos para contar con dos bases de datos ordenadas con toda la información necesaria. Como resultado obtuvimos dos archivos unificados que fueron explorados en una herramienta de Inteligencia y Datos como lo es Google Data Studio para la creación de gráficos y tablas. El conjunto de datos final con la información cuantitativa de Sistrix y sus más de 240.000 filas de datos puede ser descargado desde Google Drive[25], para el caso de Ahrefs los conjuntos de datos se encuentran en un folder de Google Drive[26] .

Análisis y exploración de datos

Cuando Project Veritas filtró una lista negra preparada por Google para que un grupo determinado de sitios web no aparezcan en Google Now (un asistente personal desarrollado por Google con multitud de funciones como recordar cumpleaños, vuelos, correos, etc. Entre estas funciones estaba la exposición de noticias[27]) hubo revuelo en el mundo periodístico norteamericano por distintos motivos, uno de los más destacados es que la existencia de esta lista contradice una clara afirmación de Google sobre su interferencia en los resultados de su buscador: “no utilizan curación humana para recolectar y organizar resultados” Sullivan, D. (2019, Julio 15). Pero más allá de la coherencia de Google entre

[24] https://www.sistrix.com/api/
[25] https://drive.google.com/file/d/1GfdpExAHkGLFA9b0CtqrCFQ8iECEGc0-/view?usp=sharing
[26] https://drive.google.com/drive/folders/14aSFt59HN0ZzjapEN_8tAfKQmaiSd9_L?usp=sharing
[27] https://es.wikipedia.org/wiki/Google_Now

sus comunicados públicos y acciones hay otras cuestiones sociales y políticas importantes; esta filtración invita a profundizar en preguntas cuya respuesta van a moldear el futuro de la información: ¿Existe una motivación política detrás de esta lista negra? Si las hay ¿cuáles son? ¿Estas motivaciones políticas son de Google, otra empresa privada o de algún lobby en particular? ¿Hasta qué punto es posible permitir que una compañía privada oculte información sin dar explicaciones? Y lo más importante: ¿Existen otras listas, modificaciones o cambios manuales que Google ejecuta sobre sus servicios y algoritmos que afecten sitios noticiosos?

Aunque no es posible dar una respuesta concreta a cada una de estas preguntas, indagamos en esta investigación con exactitud los cambios que Google hizo sobre sus algoritmos después de las elecciones de 2016 y sus implicaciones sobre los resultados de búsqueda. Después de todo una investigación interna en Google durante este año reveló que aproximadamente el 0.25% de las búsquedas posicionaron resultados con información engañosa, Gomes (2017, Abril 25), lo que llevó a Google a crear un equipo especializado en solucionar este problema, a este equipo lo llamaron internamente “Project Owl”, Grind et al. (2019, Noviembre 15). Esto en sí ya es una confirmación explícita de que Google tiene intenciones de modificar sus resultados de búsqueda con el fin de combatir la desinformación -o como muchos también pensarían, para quitar fuentes de información que no se alinean con su visión política, aunque esto no es un hecho verificable-.

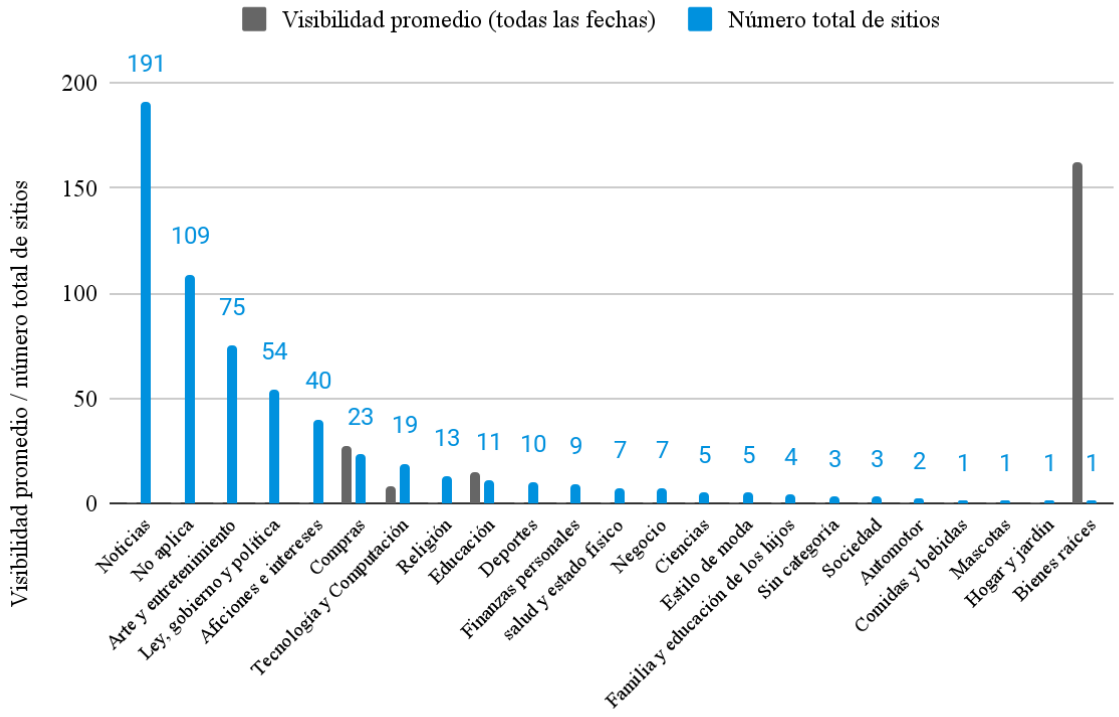
Considerando que existe una lista negra con la que Google negó la exposición de determinados sitios en Google Now, necesitamos determinar si todos los sitios, o al menos un grupo de estos sitios, experimentó cambios importantes en sus posiciones dentro de los resultados de búsqueda.

Siguiendo con esta línea de sucesos, no es extraño encontrar que entre la lista total de sitios 191 de ellos en total pertenezca a la categoría IAB Tier 1 de noticias. Esto significa que el 38.8% de los sitios presentes en la lista negra de Google están clasificados como una fuente noticiosa, un número muy destacado como para ser ignorado que resalta sobre el resto de categorías. También otras categorías importantes incluyen leyes, gobierno y política. Esto sugiere que la atención que Google pone sobre estos sitios podrían ser una respuesta a las

elecciones presidenciales de 2016 muy probablemente debido a los resultados de "Project Owl".

Figura 1

Visibilidad total y Número total de sitios IAB Tier 1

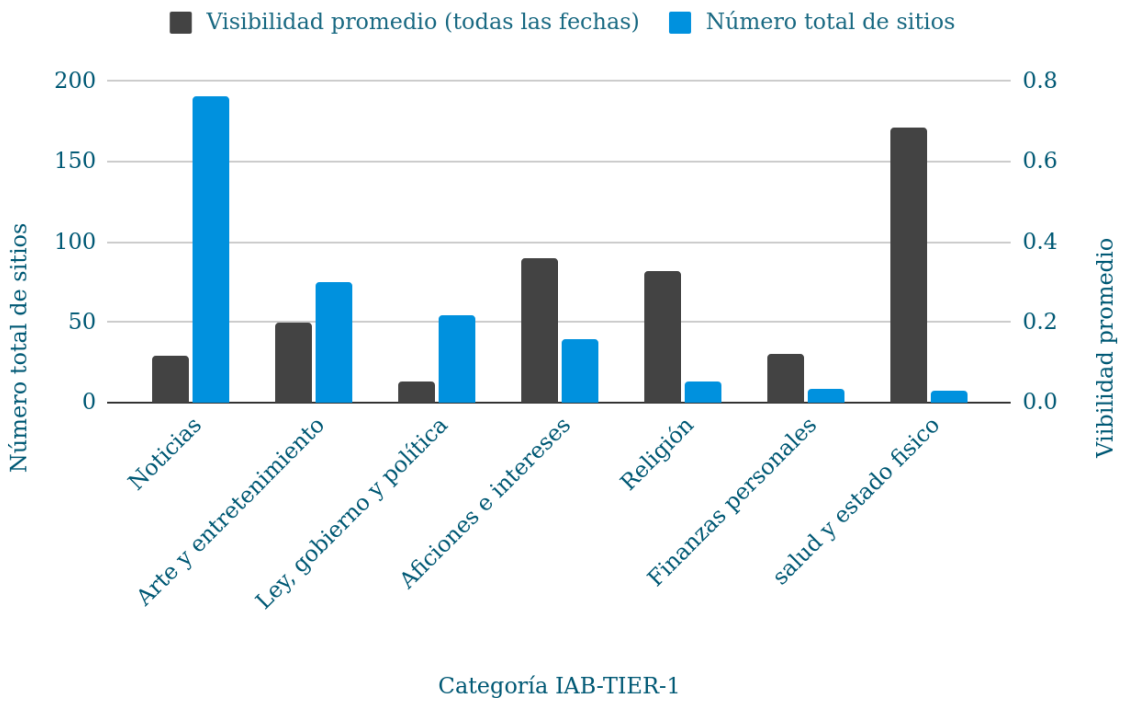


Fuente: Elaboración propia.

Con la figura 1 podemos ver cómo las categorías de Compras, tecnología y computación, educación y bienes raíces tienen una visibilidad total proporcionalmente mucho más alta que otras categorías. Esto significa que los pocos sitios pertenecientes a estas categorías son considerablemente más grandes que el promedio y que al ser además de ello un número menor de sitios podrían distorsionar algunos gráficos. Es probable que los sitios presentes en estas categorías sean excluidos de Google Now sencillamente por su bajo valor periodístico al ser netamente sitios de valor comercial. Por estos motivos a partir de ahora descartamos estas categorías de los gráficos generales lo que nos permitirá tener una visión más certera de los tipos de contenidos que verdaderamente interesan en esta investigación.

Figura 2

Visibilidad total y Número total de sitios IAB Tier 1, filtrado por categorías



Fuente: Elaboración propia.

Al acotar la Figura 1 y visualizar únicamente las categorías que tienen una potencial asociación con las *fake news* encontramos una foto mucha más clara sobre cuál es la visibilidad de cada una, en este caso las categorías de Religión, Salud y estado físco y de Aficiones e intereses son las más destacadas.

Cabe resaltar que la categoría de noticias puede presentar un nivel de visibilidad menor pero es con diferencia la categoría con más sitios en total. Además de ello algunos de los sitios contenidos en esta categoría tienen una visibilidad considerablemente superior a la media -1.7- como puede ser naturalnews.com con 4.6, torrentfreak.com con 3 y dailycaller.com con 2.1. Esto significa que la categoría de Noticias pese a tener una visibilidad promedio baja remarca una relevancia total en el estudio por su alta participación y además por contar con dominios de visibilidad alta.

Es importante aclarar que en cuanto a los sitios que pertenecen a la categoría listada como “No aplica” son sitios cuyo contenido no pudo ser agrupado dentro de una de las categorías estándares del IAB. Por lo que se debe entender a estos sitios como contenido no convencional como puede ser contenido que infringe derechos de autor, pornografía, discurso de odio, etc. Según Zvelo estos sitios estarían categorizados en estas verticales:

Figura 3

Número total de sitios según la categorización de Zvelo para sitios "No aplicables" según IAB Tier 1

Fuente: Elaboración propia.

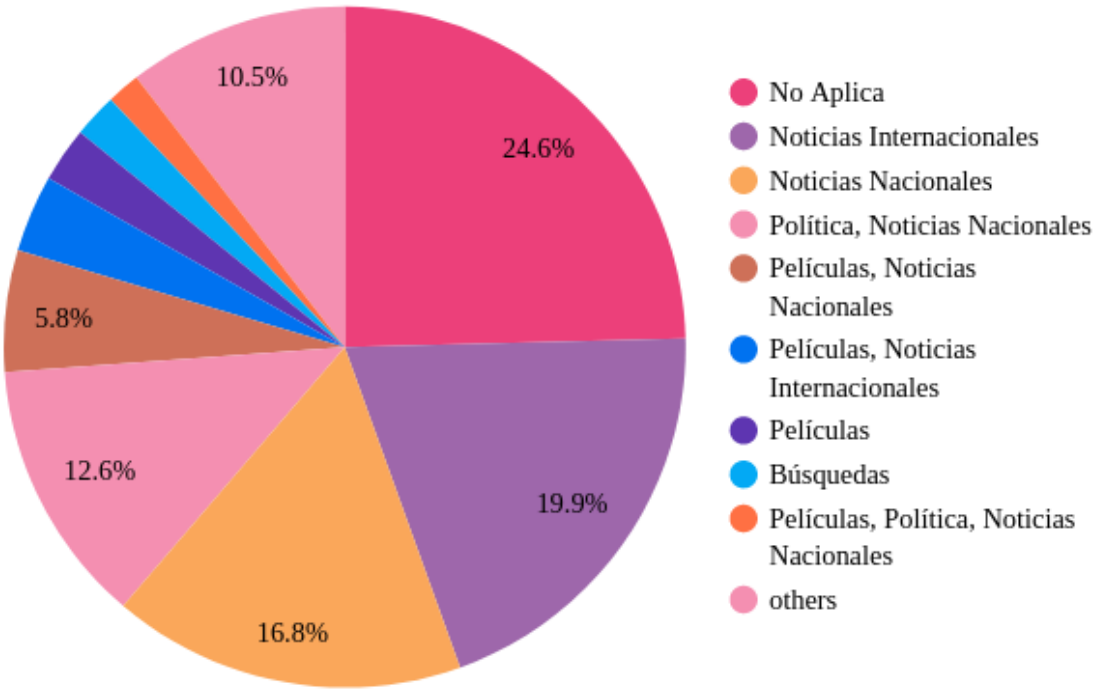
Como podemos observar, la mayoría de estos sitios que se encuentran dentro de la categoría “no aplica” -la segunda categoría más grande- corresponden a contenido de blogs y foros comunitarios, sin embargo la existencia de contenido que podríamos catalogar como “problemático” es notorio -discurso de odio, pornografía, armas, entre otros-.

Esta agrupación a nivel general expone de forma sutil el objetivo de la lista negra: disminuir la exposición de determinados sitios considerados como problemáticos, o bien, sitios que no sean importantes a nivel noticioso.

Lo interesante y valioso de este conjunto de datos es la posibilidad de indagar en detalle sobre las distintas subcategorías, en este caso a través de la Figura 4 podemos ver cómo se componen las subcategorías de la categoría IAB Tier 1 de "Noticias".

Figura 4

Distribución porcentual de las subcategorías de Noticias según el IAB Tier 2



Fuente: Elaboración propia.

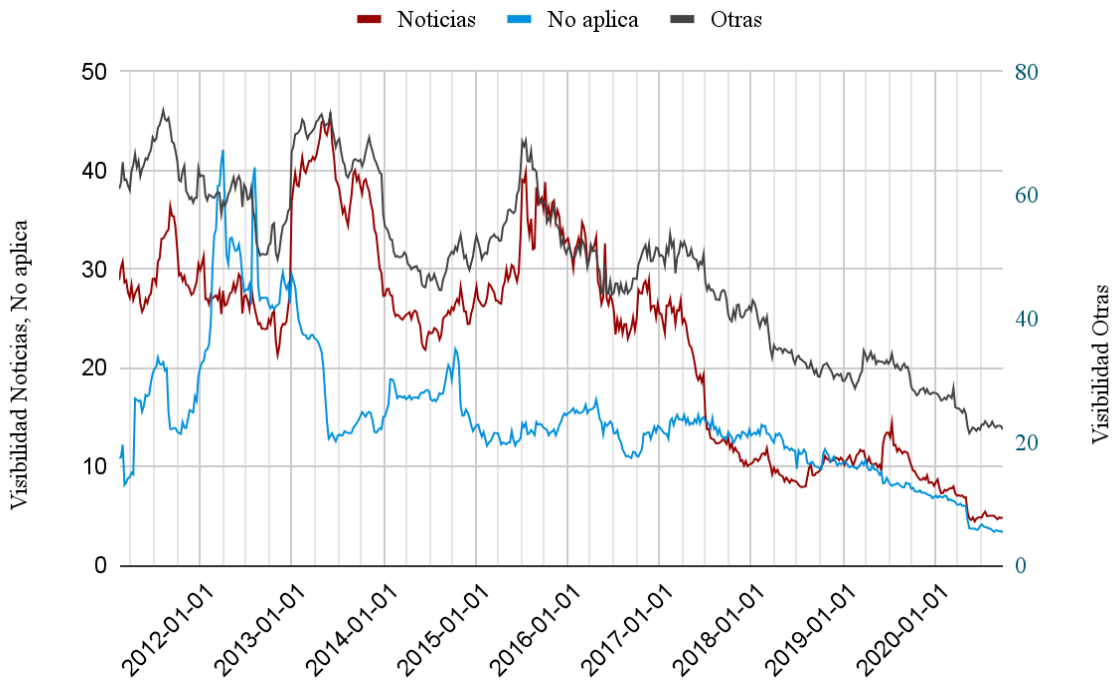
Como se puede ver incluso en las subcategorías pueden existir genéricos como el No Aplica, es decir, sitios de noticias genéricos que no obedecen a una subcategoría específicos, también llama la atención la alta presencia de sitios de noticias "nacionales" e "internacionales". Según esto la gran mayoría de sitios de noticias incluidos en esta lista no abarcan una temática específica. Esto implica que para entender su comportamiento es necesario evaluar su contenido desde otro punto de vista. También resaltamos la presencia

de la etiqueta “Películas” en varios sitios de noticias abriendo la posibilidad de que varios de estos sitios de noticias estén asociados a contenido de piratería.

Volviendo a la categorización IAB Tier 1, al observar que las categorías mayoritarias en número total de sitios son “Noticias” y “No aplica” optamos por realizar un histograma en donde sumamos los puntos de visibilidad por día de todos los sitios que comparten esa categoría en común. En este caso vemos que la tendencia específica de la categoría “Noticias”, aunque comparte una tendencia general de caída de visibilidad sí tiene una tendencia de caída notoria a mediados de 2017.

Figura 5

Visibilidad por categorías según la clasificación IAB Tier 1



Fuente: Elaboración propia.

Como puede observarse al agrupar todas las categorías -excluyendo Compras, Tecnología y Computación, Educación y Bienes raíces- tienen una tendencia a la baja a partir de junio de 2017, pero resalta que “Noticias” experimente una caída de visibilidad mucho más marcada y abrupta que las demás, mientras que las otras experimentan una tendencia a la baja relativamente estable. En el gráfico resaltamos una clara tendencia negativa entre finales de abril y principios de julio de 2017 -siendo especialmente llamativo el día 25 de junio al ser

la fecha donde se observa la caída más pronunciada-. Este tipo de tendencias tan abruptas tienden a ocurrir a causa de una actualización de los algoritmos de Google. Para analizar en detalle este comportamiento creamos una tabla con los cambios porcentuales de la visibilidad de cada categoría comparando dos periodos, un año antes y un año después del 25 de junio.

Tabla 3

Comparativa de categorías de sitios agrupados por IAB Tier 1.

IAB-TIER-1 Traducción	Visibilidad promedio (todas las fechas)	Número total de sitios	Visibilidad promedio 2016-06-25 - 2017-06-25	Visibilidad promedio 2017-06-26 - 2018-06-26	Cambio absoluto	Cambio porcentual
Noticias	0.119	191	0.130	0.058	-0.071	-55.08%
No aplica	0.137	109	0.123	0.121	-0.002	-1.74%
Arte y entretenimiento	0.199	75	0.197	0.141	-0.056	-28.56%
Ley, gobierno y política	0.053	54	0.047	0.017	-0.030	-63.74%
Aficiones e intereses	0.358	40	0.463	0.462	-0.001	-0.29%
Compras	27.607	23	30.497	35.491	4.994	16.38%
Tecnología y Computación	8.418	19	12.453	12.875	0.423	3.39%
Religión	0.326	13	0.303	0.195	-0.108	-35.61%
Educación	15.076	11	18.945	24.242	5.297	27.96%

IAB-TIER-1 Traducción	Visibilidad promedio (todas las fechas)	Número total de sitios	Visibilidad promedio 2016-06-25 - 2017-06-25	Visibilidad promedio 2017-06-26 - 2018-06-26	Cambio absoluto	Cambio porcentual
Deportes	0.045	10	0.036	0.067	0.032	88.17%
Finanzas personales	0.121	9	0.121	0.203	0.083	68.46%
salud y estado físico	0.684	7	0.335	0.072	-0.263	-78.56%
Negocio	0.160	7	0.139	0.153	0.014	10.19%
Ciencias	0.013	5	0.022	0.006	-0.016	-71.95%
Estilo de moda	0.008	5	0.008	0.006	-0.002	-22.73%
Familia y educación de los hijos	0.287	4	0.227	0.379	0.151	66.52%
Sin categoría	0.000	3	0.000	0.000	0.000	0.00%
Sociedad	0.208	3	0.267	0.303	0.036	13.45%
Automotor	0.006	2	0.013	0.008	-0.005	-38.65%
Comidas y bebidas	0.038	1	0.040	0.027	-0.013	-32.74%
Mascotas	0.001	1	0.006	0.001	-0.005	-89.33%
Hogar y jardín	0.001	1	0.005	0.003	-0.002	-43.16%

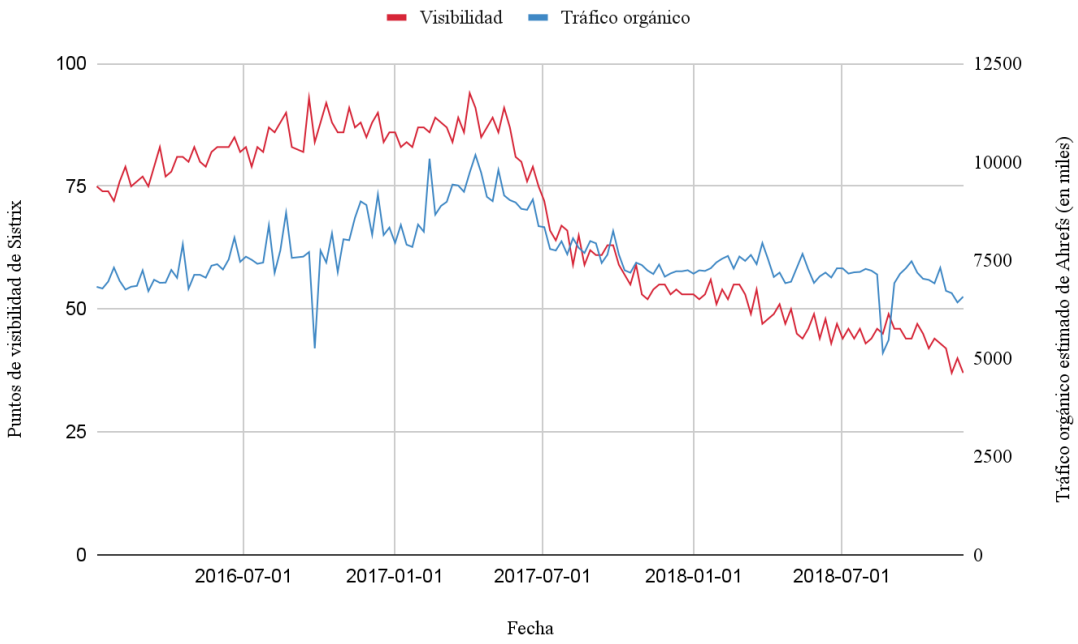
IAB-TIER-1 Traducción	Visibilidad promedio (todas las fechas)	Número total de sitios	Visibilidad promedio 2016-06-25 - 2017-06-25	Visibilidad promedio 2017-06-26 - 2018-06-26	Cambio absoluto	Cambio porcentual
Bienes raíces	161.933	1	202.352	258.880	56.528	27.94%

Fuente: Elaboración propia.

Es notorio que en cuanto a valores porcentuales y absolutos la pérdida de visibilidad es mucho más notoria en los sitios relacionados con noticias, política, religión, salud y estado físico, arte y entretenimiento. Aunque hay otras categorías con incrementos o caídas en visibilidad mucho más notorias a nivel porcentual, son cambios que no son significativos al ser categorías con pocos sitios o con poca visibilidad absoluta.

Figura 6

Visibilidad y tráfico orgánic de sitios según la clasificiación de noticias IAB Tier 1



Fuente: Elaboración propia.

Al hacer un estudio en detalle sobre la categoría de noticias, vemos que los datos de Ahrefs y Sistrix coinciden en que existe una caída notoria y constante para los sitios de esta lista a partir de mediados de 2017.

Encontramos que hay dos fechas clave: 17 de abril y 25 de junio de 2017. Justo cuando distintas comunidades especializadas en motores de búsqueda exponen sospechas de que Google ha podido publicar actualizaciones. Schwartz, B. (2017, Abril 19), un reconocido autor en la industria del SEO -optimización de posicionamiento en motores de búsqueda- comenta en su artículo sobre la posible existencia de una actualización de Google justo el 17 de abril basándose en los testimonios de distintos usuarios y, además, en los cambios en el posicionamiento reportados por distintas herramientas. Ocurre lo mismo con el 25 de junio, Schwartz, B. (2017, Junio 27) explica bajo argumentos y pruebas similares la existencia de una actualización de Google a partir de esta fecha.

Lo importante de esto está precisamente en el desconocimiento general sobre estas dos actualizaciones detectadas. Por lo general cada actualización se notifica mediante un comunicado por parte de Google explicando los propósitos de su actualización. O, como mínimo, la comunidad profesional SEO se encarga de exponer hipótesis precisas basadas en datos y experiencias comunitarias. En este caso no ocurre ninguno de las dos premisas citadas: ni existe un comunicado por parte de Google ni encontramos explicaciones precisas por parte de un tercero. Lo que sí encontramos son explicaciones hipotéticas como las de Rank Ranger[28] que, de todas formas, no logran ser precisas.

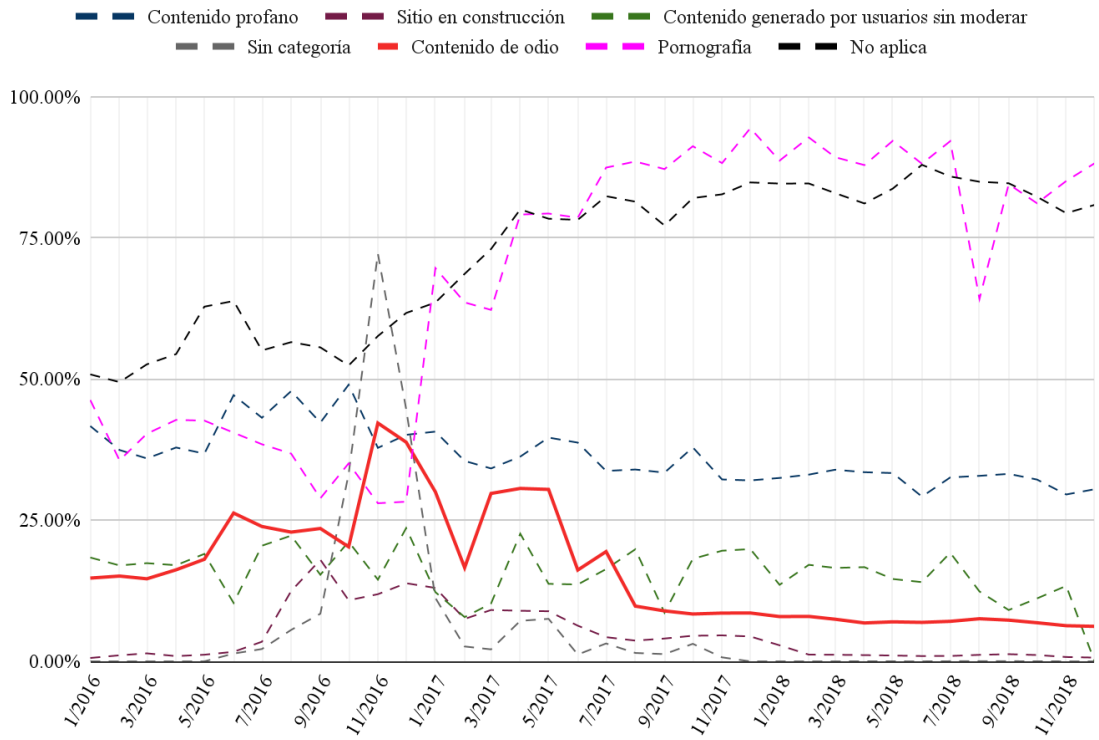
Los datos presentes en esta investigación parecen aclarar por primera vez que estas actualizaciones de 2017 tuvieron como objetivo disminuir el posicionamiento de sitios noticiosos que no estuvieran alineados con determinados principios -que podríamos catalogar presuntamente como veracidad, integridad y confiabilidad-. Esto es muy importante, porque supone el hallazgo del primer algoritmo de Google destinado a combatir la desinformación en su historia.

Para verificar esto creamos un gráfico utilizando la estimación de tráfico orgánico de Ahrefs en números relativos según la categorización de IAB non standard. Con esto notamos un detalle importante: solo los sitios categorizados como contenido de odio eran los que tenían una caída significativa en la estimación porcentual del tráfico orgánico.

[28] https://www.rankranger.com/blog/google-algorithm-update-june-2017-explained

Figura 7

Visibilidad y tráfico orgánic de sitios según la clasificiación de noticias IAB Tier 1



Fuente: Elaboración propia. Promedio mensual del porcentaje de tráfico orgánico por categoría, calculado a partir de datos diarios (tráfico total estimado del día / tráfico total estimado máximo histórico).

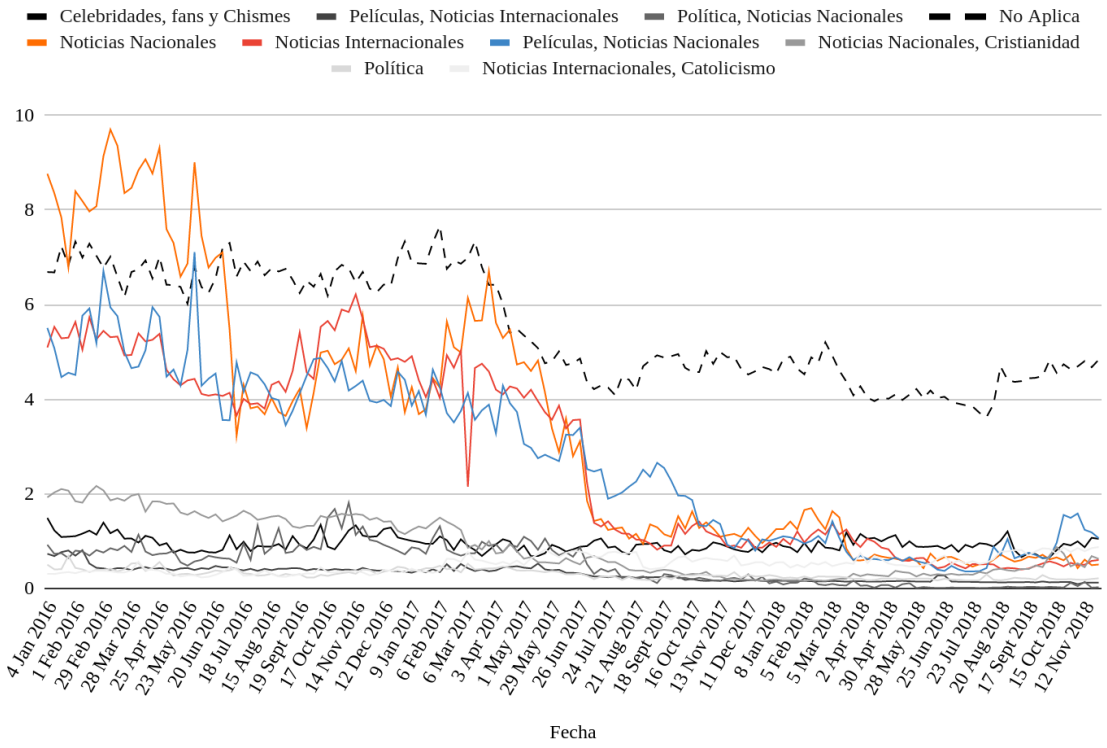
Es evidente que sobre las fechas clave -abril y junio de 2017- el contenido de odio ha sido el único con una tendencia claramente negativa en la estimación de tráfico orgánico de Ahrefs. De hecho, otras categorías problemáticas como la pornografía aumentan y el contenido profano -es decir, contenido profundamente ofensivo y soez- es relativamente

estable pese a tener una tendencia a la baja. Es evidente que las actualizaciones de Google impactaron en este caso a un grupo específico de sitios.

Los datos vistos en este análisis evidencian un cambio de Google que afecta a los sitios de noticias, pero además de ello apunta específicamente a sitios problemáticos -contenido de odio-; sin embargo, no queda del todo claro cómo estos cambios afectan a los sitios de noticias, ¿son determinados sitios de noticias? ¿Determinadas subcategorías? ¿Algunos grupos específicos?

Figura 8

Evolución de visibilidad para los sitios de noticias en IAB Tier 1 divididos por su correspondiente categorización en IAB Tier 2



Fuente: Elaboración propia.

Cuando evaluamos la caída de la visibilidad de los sitios de noticias dividiéndolos según su categoría IAB Tier 2 encontramos que, aunque todas las categorías caen, en términos generales las más afectadas son Noticias Nacionales y Noticias Internacionales.

Considerando que no se ha encontrado que los sitios de noticias tengan agrupaciones notorias en categorías problemáticas, intentamos organizar los sitios que más perdieron visibilidad.

Tabla 4

Top 20 de dominios organizados por visibilidad promedio con su respectivo cambio absoluto y porcentual de visibilidad.

Sitios de Noticias	Visibilidad promedio (todas las fechas)	Visibilidad promedio 2016-06-25 - 2017-06-25	Visibilidad promedio 2017-06-26 - 2018-06-26	Cambio absoluto	Cambio porcentual
naturalnews.com	4.688	2.237	0.352	-1.885	-84.25%
torrentfreak.com	3.036	3.580	2.863	-0.717	-20.03%
dailycaller.com	2.135	3.104	1.203	-1.901	-61.24%
christianpost.com	1.676	1.206	0.306	-0.900	-74.62%
infowars.com	1.189	1.411	0.478	-0.933	-66.13%
opposingviews.com	0.692	0.127	0.007	-0.119	-94.30%
advocate.com	0.683	1.004	0.504	-0.499	-49.74%
mediamatters.org	0.581	0.729	0.237	-0.492	-67.45%
americanthinker.com	0.510	0.468	0.065	-0.402	-86.06%
catholicnewsagency.com	0.462	0.452	0.580	0.128	28.24%
redstate.com	0.459	0.812	0.179	-0.633	-78.00%
newsbusters.org	0.420	0.321	0.075	-0.246	-76.75%
coolthings.com	0.368	0.528	0.612	0.084	15.99%
frontpagemag.com	0.364	0.360	0.235	-0.126	-34.86%
glennbeck.com	0.353	0.218	0.117	-0.101	-46.16%
out.com	0.342	0.482	0.452	-0.030	-6.28%
twitchy.com	0.261	0.385	0.124	-0.261	-67.74%
drudge.com	0.232	0.232	0.132	-0.099	-42.87%

lifenews.com	0.229	0.273	0.060	-0.213	-78.00%
thegatewaypundit.com	0.218	0.637	0.059	-0.578	-90.66%

Fuente: Elaboración propia.

La tabla 4 nos muestra un patrón de comportamiento bastante claro casi todos los sitios categorizados en "Noticias" han perdido más del 30% de su visibilidad tras la actualización de Google durante 2017. En concreto el 86,38% de los sitios web dentro de esta categoría perdieron por lo menos el 30% de su visibilidad. En total todos los sitios de esta categoría tienen una caída del 56% de su visibilidad. No obstante, ninguno de los sitios presentes en la tabla 4 fue catalogado como contenido de odio y solo 3 de ellos son etiquetados como contenido "no seguro para marcas". Esto nos lleva a preguntarnos qué tienen en común y cómo se explica su pérdida de posicionamiento en Google.

Lo que tienen en común todos estos sitios -tras verificar de forma manual cada uno de ellos- es que tienen una alta presencia de contenido político. Esto es sin duda un punto de partida para futuras investigaciones, en donde aparentemente los sitios de noticias que tratan específicamente temas políticos -o al menos mayoritariamente políticos- fueron los más afectados tras las actualizaciones de Google en 2017. Este hecho no solo es interesante para una investigación orientada al análisis de contenidos, sino que también responde a lo que presuntamente sería uno de los objetivos de Google tras las elecciones de 2016: impedir la exposición de contenido político considerado como malicioso o dañino.

Todo esto implica que Google ha modificado los algoritmos asociados a su motor de búsqueda para limitar la aparición de determinados sitios basándose en su mensaje político, o cuanto menos, basándose en la veracidad de sus noticias. El problema que esto conlleva es que hasta donde sabemos no hay un programa de verificación de datos tras estos cambios y tampoco existe un comunicado de Google anterior a junio de 2017 distintas al código de datos estructurados que puso en disponibilidad en 2016[29]. Por tanto, a no ser que nueva evidencia indique lo contrario, podemos decir que Google en 2017 alteró el posicionamiento dentro de su buscador basándose en decisiones propias y no en verificaciones de datos externas.

[29] https://www.theguardian.com/technology/2016/oct/13/google-news-fact-check-trump-clinton-us-election

Otra posibilidad es que Google de forma inocente actualizó su algoritmo para favorecer el posicionamiento de entidades -sitios, marcas y empresas- con mayor reconocimiento. Por ejemplo: "la *CNN* es un medio noticioso más reconocido que naturalnews.com; por tanto, *CNN* debería tener un mayor posicionamiento en Google para cualquier contenido en comparación a naturalnews.com". Esto también podría explicar lo sucedido y haría que la pérdida de visibilidad de los sitios de noticias estudiados en este artículo no sea más que un producto de la casualidad. De todas formas, lo que es un hecho es que estos sitios fueron excluidos de forma manual de Google Now, lo cual no hace extraño que las actualizaciones de los algoritmos también fueran modificados para impedir su posicionamiento.

En lo que coincidimos es que, tras revisar los dominios de noticias, notamos que todos en términos generales tienen un bajo nivel de calidad y la veracidad de los hechos publicados en ellos son en muchos casos cuestionables, incluso en muchos de ellos sospechamos de que sus textos son orientados a la manipulación de la audiencia. Verificar este hecho requeriría una nueva investigación, pero nos abre la posibilidad de investigar sobre cómo Google puede reconocer la calidad de un sitio noticioso. Después de todo, un ser humano no puede verificar todo el contenido disponible. Los resultados del algoritmo tienen que provenir de un proceso escalable.

Explicación e indagación teórica

Como muestran distintas referencias y además evidencian los resultados cuantitativos, Google tuvo una participación activa en los cambios de posicionamiento para distintos tipos de noticias. La explicación oficial a estos resultados está en el artículo de Gomes (2017, Abril 25) quien menciona que Google realizó cambios en las señales que reciben sus algoritmos para promover sitios con mayor autoridad mientras se degradan sitios con contenido de baja calidad. Esta explicación cuenta con dos problemas, el primero es que no cuenta con ningún detalle relevante ya que no especifica cómo Google determina si un sitio web tiene un alto nivel de autoridad, y por otro lado tampoco explica a qué tipo de contenido de baja calidad se refiere ya que los resultados observables en nuestro estudio sugieren que la motivación no se centra específicamente en el "contenido de baja calidad" sino en el contenido noticioso y político. Además, el hecho de que muchos de estos sitios sean añadidos en una lista de forma manual sugiere que las modificaciones no dependen

exclusivamente de una actualización en los algoritmos; por el contrario, existe una intervención humana.

Esto fue señalado por Grind et al. (2019, Noviembre 15), quienes afirman que precisamente debido a la presencia de contenido desinformativo la intervención humana en los resultados de búsqueda era un aspecto rutinario dentro de Google, pero al mismo tiempo Google nunca realizó una declaración pública en donde detalle dichas intervenciones ni las motivaciones detrás de las mismas. Estamos frente a un problema de opacidad en donde lo que se nos es distribuído ocurre por decisión de una compañía privada.

En este mismo artículo también mencionan la lista negra de Google -el objeto de estudio de este artículo- de una forma muy particular:

> "La práctica de crear listas negras para ciertos tipos de sitios o búsquedas ha alimentado gritos de sesgo político de algunos ingenieros de Google y publicaciones de derecha que dijeron haber visto partes de las listas negras. Algunos de los sitios web que Google parece haber apuntado en Google News eran sitios y blogs conservadores, según documentos revisados por el diario. En una lista negra parcial revisada por el diario, algunos sitios web conservadores y de derecha, incluidos The Gateway Pundit y The United West, se incluyeron en una lista de cientos de sitios web que no aparecerían en noticias o productos destacados, aunque podrían aparecer. en los resultados de búsqueda orgánicos." Grind et al. (2019, Noviembre 15)

Vale la pena destacar que The Gateway Pundit (thegatewaypundit.com) aparece en la lista estudiada en este artículo, y este dominio también vio reducida sus posiciones en la búsqueda orgánica de Google. Esto significa que estas listas manuales también tienen una implicación -sea directa o indirecta- sobre las posiciones orgánicas de un sitio web.

Conclusiones y propuestas investigativas

Los resultados cuantitativos en contraste con los hechos ocurridos en 2016 y 2017 -las elecciones norteamericanas y las actualizaciones en los algoritmos de Google- dejan en evidencia una reacción de Google frente al resultado de las elecciones. Todo apunta a que Google realizó modificaciones manuales con el fin de reducir la visibilidad de un grupo

específico de sitios. En este sentido es posible que las acciones de Google tengan una repercusión positiva, pero esto requiere de una posterior investigación en la que se necesita de un análisis de contenidos y una verificación de hechos en las historias presentes en los sitios noticiosos expuestos en la lista de esta investigación.

Los resultados respaldan -o por lo menos sugieren- que las sospechas de otros autores sobre la forma en la que Google opera su buscador son ciertas: hay una intervención humana sobre los resultados de búsqueda de una forma directa o indirecta. Cambios en el posicionamiento de sitios como The Gateway Pundit no son casualidad. Así mismo los resultados exponen cambios algorítmicos que apuntan a sitios de noticias, en donde evidentemente busca un cambio en el posicionamiento de estos.

Más allá de los hechos, debemos partir de que estos cambios e intervenciones tienen buenas intenciones, pero la opacidad de Google al respecto es cuanto menos preocupante para el conjunto de la sociedad, ya que el flujo libre de información y la verificación de la misma no debería recaer únicamente en la decisión de entidades privadas. El derecho a la información es un derecho del público y es una decisión en donde gobiernos y personas deben involucrarse. En este caso es una acción unidireccional en donde Google tiene la última palabra.

Una línea de investigación a futuro podría estar enfocada en el seguimiento del posicionamiento de un amplio espectro de sitios noticiosos de distintas orientaciones políticas incluyendo a los extremos. Un seguimiento del posicionamiento de estos sitios frente a las actualizaciones de Google podría ayudar a establecer si las intervenciones humanas de Google han dado lugar a decisiones parciales y poco benéficas para la neutralidad de la información.

Por el momento el debate sobre las *fake news* sigue abierto, e indiscutiblemente Google es una de las compañías que tiene el control indiscutible sobre gran parte de su distribución.

MALWARE Y *FAKE NEWS*: UN NUEVO FRENTE DE TRABAJO PARA COMBATIR LA DESINFORMACIÓN

Introducción

Para comprender la importancia que supone esta línea investigativa es necesario imaginar un escenario hipotético: un grupo de hackers ha logrado vulnerar la seguridad de miles de dispositivos móviles y de escritorio en España a través de un software malicioso dedicado a modificar la información que los usuarios leen en Internet. Con ello han conseguido que siempre que los usuarios propietarios de dichos dispositivos consuman la información de cualquier periódico reconocido en España (como El Mundo, El Confidencial, El País u otro similar) vean una *Fake New* en la página principal de cada uno de estos periódicos. Al ser algo que ocurre en el dispositivo del usuario ninguno de estos periódicos puede hacer algo al respecto -no pueden impedir de ninguna forma que los usuarios consuman y vean estas *fake news* en sus sitios web-, y al ser un software oculto ninguno de los usuarios puede darse cuenta de lo que sucede. Ni los periódicos ni los usuarios son conscientes de la masiva distribución de información falsa que ocurriría en este escenario lo que garantiza un impacto estruendoso de la desinformación en la población además de ser un duro golpe para la credibilidad de todos los periódicos del país, en otras palabras: un cambio irreversible en la sociedad española.

Todos podemos coincidir en que la desinformación transmitida en este escenario hipotético sería aún más peligrosa que las ya conocidas ya que en este caso cada *Fake New* se encuentra "insertada" en un medio de comunicación creíble y confiable haciendo que el material informativo tradicional de educación frente a las *fake news* sea de facto obsoleto ya que "leer noticias de fuentes confiables" sería una recomendación completamente inocua en este escenario, después de todo es precisamente la *Fake New* la que aparece insertada dentro de los periódicos más confiables de España en este escenario hipotético.

Dicho esto, es necesario preguntarse si esto sería posible, y de serlo se tiene que pensar en cómo sería posible. En esta investigación se parte de la hipótesis de que la realidad es bastante sencilla: sería perfectamente posible que esto sucediera, y sería posible haciendo uso de software malintencionado -*Malware*-. Por ello en esta investigación se expone

conceptualmente cómo sería posible lograr dicho escenario hipotético bajo herramientas y técnicas de distribución de software actualmente disponibles.

Lo verdaderamente importante está en reconocer que este escenario significa un cambio en el paradigma de la distribución del contenido malicioso: ahora no basta con la concienciación de la población para que confíen en fuentes sólidas, tampoco basta con crear sofisticados sistemas de detección de noticias falsas para reducir su distribución a través de redes sociales; también es necesaria la educación frente a la seguridad informática y una eficiente persecución del software malicioso cuyo fin sería la distribución de noticias falsas.

Esta línea investigativa deja claro que la distribución de la información falsa no depende solo de Meta, Google, WhatsApp, Telegram o similares, sino que el dispositivo (tablet, laptop, teléfono móvil, etc.), el sistema operativo (Windows, Linux, Android, MacOS, etc.), los navegadores (Firefox, Chrome, Safari, etc.) y otro software/hardware involucrado tiene el potencial de ser usado para promover la distribución de noticias falsas de formas aún más agresivas y efectivas, todo a través de una escalada tecnológica que es inevitable: la información dañina siempre encontrará nuevos medios de distribución y el uso del software malicioso simplemente es un paso lógico en este proceso, incluso si no se ha evidenciado su uso en la actualidad. Si puede ser usado, será utilizado.

Tanto la seguridad informática como las *fake news* son temas investigativos demasiado amplios y dado que nuestro objetivo es exponer una prueba de concepto es necesario acotar en detalle qué es lo que abarca este proceso: una prueba de concepto que consiste en crear un software malicioso que funciona en el cliente -el dispositivo del usuario- cuyo fin es modificar la información presente en un sitio de noticias confiable para que incluya contenido desinformativo. Es decir, la distribución de las noticias falsas no va a depender en ningún caso de un sitio web, una página en Facebook, un perfil en Twitter o de algún otro medio de acceso público a la información, sino que será un software instalado en el navegador del usuario el que estará encargado de insertar y exponer *fake news* al usuario afectado.

Por tanto, nuestra investigación excluye una amplia gama de conceptos pertenecientes a la seguridad informática como el estudio de vulnerabilidades, técnicas de desarrollo, vectores de ataque, etc. En cuanto a las *fake news* solo nos limitamos a demostrar su posible

distribución, pero no abarcamos temas adicionales como el análisis de contenido, su impacto en la sociedad, su detección, etc. El objetivo es demostrar que es posible distribuir contenido desinformativo haciendo uso de software malicioso en un escenario que a día de hoy es posible.

Metodología

Como bien se ha dicho, el objetivo de este capítulo es demostrar que es posible distribuir *fake news* usando *Malware* o software malicioso. Para ello se ha optado por crear una "prueba de concepto", es decir, crear un escenario en donde se ejemplifica de forma evidente la distribución de información falsa.

Para crear esta prueba de concepto y demostrar el potencial actual existente para distribuir noticias falsas mediante software malicioso -siendo este un vector de distribución mucho más complejo de controlar y de detectar- es necesario tener en cuenta que existen muchísimos tipos de *Malware* y a su vez existen diversos métodos de ataque, lo cual hace que sea necesario elegir un método de ataque concreto en un escenario específico dado que las posibilidades técnicas de distribuir y ejecutar software en un dispositivo son demasiado extensas. Intentar crear una prueba de concepto que abarque distintas tecnologías, dispositivos y métodos de distribución de contenidos sería inabarcable.

Por ello la metodología está centrada en crear un caso concreto que pueda ser replicado en un escenario real de una forma plausible, es decir, por un lado no está planteado un escenario demasiado específico o difícil de replicar por algún actor malintencionado y por otro se busca que dicho escenario sea centrado en una única tecnología y plataforma; algo fácil de replicar en una tecnología concreta.

Es importante aclarar que el objetivo no es demostrar la existencia de una vulnerabilidad informática sino exponer una prueba de concepto en la que el software malicioso es un potencial distribuidor de *fake news,* esto implica que el resultado de este experimento no va a desembocar en una actualización de seguridad ni tampoco reflejar una mala práctica de seguridad informática; por el contrario, el resultado solo refleja la posibilidad de usar software para distribuir *fake news*. Por ello es más importante encontrar un método simple y efectivo por encima de un método técnicamente complejo o incluso una vulnerabilidad

informática importante. Esto además debe tomarse en cuenta como punto de evaluación para los resultados de esta investigación, ya que el valor de este escenario hipotético no está en la complejidad aplicada desde la perspectiva de las ciencias informáticas, o en un ingenioso ataque que explote alguna vulnerabilidad desconocida; el valor está precisamente en lo contrario: demostrar que sin una alta complejidad técnica y sin depender de una vulnerabilidad en lo que respecta a la seguridad informática es posible conseguir un efecto dañino de alto impacto a través de las *fake news*.

Para conseguir lo anterior lo primero es limitar la función del software que será usado en esta prueba, por lo cual metodológicamente todo está limitado a un software que engañe al usuario ofreciendo una característica deseada pero que a su vez ejecuta procesos indeseados sin que el usuario sea notificado de forma alguna, en otras palabras: prometer una función y hacer otra completamente diferente de forma inadvertida para cualquier usuario. Lo segundo es definir el alcance, el cual es publicar este software en la tienda *Chrome Web Store*.

Que sea una aplicación publicada en la *Chrome Web Store* es importante porque cualquier aplicación aprobada en esta tienda podrá instalarse en el navegador *Google Chrome* -o cualquier otro basado en *Chromium* como Edge o Brave- del usuario sin mayores complicaciones y sin requerir amplios conocimientos informáticos, es decir: cualquier persona con la capacidad de usar un ordenador debería tener la capacidad de instalar y usar este software por su propia cuenta. Esto cubre en gran medida una de las grandes necesidades de esta prueba de concepto que es la de crear un escenario con un gran alcance y de fácil penetración en la población en general.

Es importante considerar que este software malicioso tendrá que tener como base hipotética de su distribución la necesidad de que sea instalado de forma voluntaria por el usuario al no explotar ninguna vulnerabilidad técnica, en otras palabras: estamos creando un software cuya distribución depende del engaño a los usuarios.

El uso de *Google Chrome* es perfecto debido a que es el navegador web más popular -con una cuota superior al 60% del mercado global[30]- además es muy fácil para un usuario instalar una extensión en este navegador -solo requiere un simple clic y no exige mayores

[30] https://www.statista.com/statistics/544400/market-share-of-internet-browsers-desktop/

privilegios administrativos en el sistema-. Así mismo estas extensiones ya constituyen un método de distribución actual de *Malware*[31] lo que hace factible que la prueba de concepto expuesta en esta investigación ocurra en un futuro próximo, o al menos hace factible el escenario hipotético planteado en la introducción.

Por otra parte, limitar la prueba al uso de extensiones de *Google Chrome* permite demostrar que con técnicas de baja complejidad es posible distribuir *fake news* de una forma efectiva que es invisible al usuario y a los publicadores de noticias.

Para cumplir con lo anterior metodológicamente se crea y publica un *Malware* que ataca al dispositivo del usuario -que de ahora en adelante llamaremos "cliente"- y no a los periódicos directamente -es decir, a los servidores que se encargan de que las webs de cada periódico funcionen adecuadamente- en donde dicho software será instalado de forma voluntaria por el mismo usuario.

Con lo anterior se está cumpliendo con unos aspectos mínimos necesarios para que nuestra prueba de concepto pueda ser ejecutada: usamos un software de baja complejidad -cualquier programador malintencionado podría hacer lo mismo- con la posibilidad de lograr una amplia distribución -millones de usuarios podrían ser víctimas de este método- usando un método de ataque simple -no es detectado por ningún antivirus y los usuarios pueden instalar el software fácilmente-.

El software diseñado se llama "Sonido de lluvia personalizado" y su promesa funcional es la de otorgar un "Sonido de lluvia personalizado basado en las noticias que lees". Esta aplicación es basada en el código original de Joaquín Vicente[32] y es modificada para inyectar información falsa en el sitio web de *La Vanguardia* (lavanguardia.com) a través de una modificación del *Document Object Module* (DOM). Gracias a autores como Zaini & Zainal (2018) sabemos que existe *Malware* que utiliza esta misma metodología técnica, lo que nuevamente expone la posibilidad de que este escenario hipotético sea real en el futuro.

Se ha limitado este experimento a *La Vanguardia* por dos motivos: 1. demostrar que la distribución de noticias falsas puede ser un ataque dirigido a sitios web específicos, tales como portales de alto reconocimiento y confianza (facilitando la distribución y creencia

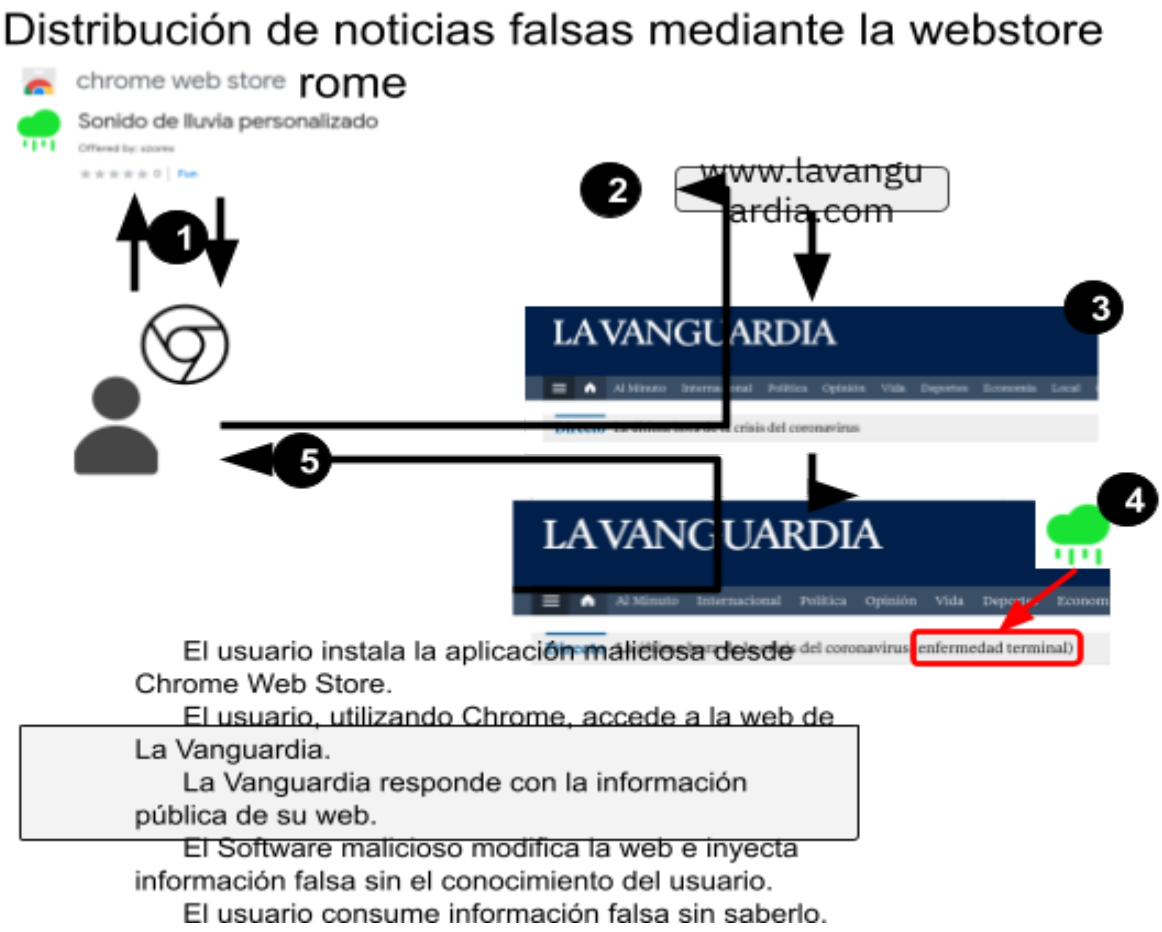
[31] https://www.tomsguide.com/us/chrome-extension-security-problems,news-26082.html
[32] https://github.com/wacko/rainsound

sobre la noticia falsa). 2. *La Vanguardia* es uno de los periódicos más leídos según el AIMC EGM (2020) en sus resultados de primera ola, de forma que si este Malware es aprobado por Google y es publicado en su tienda de aplicaciones demostrará que es posible distribuir *fake news* bajo este método y que distintos portales web no cuentan con las medidas de seguridad necesarias para evitar que esto ocurra.

El fin último de esta aplicación es la de prometer a los usuarios una función "deseable", mientras que sin que puedan percibirlo la información noticiosa que lean sobre el Coronavirus en *La Vanguardia* será modificada arbitrariamente sin notificación alguna. Esto, en otras palabras, implica distribuir una *fake news* a través de su inserción directamente en los portales noticiosos. Algo con mucho más impacto y mucho más difícil de controlar que otros métodos de distribución de la desinformación. Ya existe *Malware* diseñado para modificar el contenido de una página web a través de extensiones de navegadores, Urban (2018); por lo que una distribución de *fake news* utilizando esta técnica es algo posible, además de demostrar que nuestra metodología refleja un escenario factible. De hecho, Jagpal *et al* (2015) mencionan que uno de los métodos más comunes usados por el *Malware* distribuido a través la Chrome Web Store es la reescritura del DOM -la modificación del contenido de una página web-.

Figura 1

Diagrama de funcionamiento de la extensión maliciosa de Google Chrome



Fuente: elaboración propia.

Dicho esto, el experimento consta de 4 etapas.

1. Diseño de un software malicioso en forma de extensión para Google Chrome cuyo fin es inyectar información falsa en lavanguardia.com.
2. Publicación del software malicioso en la tienda de extensiones de Google Chrome para evaluar si este software es aceptado o rechazado.
3. Exponer el funcionamiento del software malicioso, demostrar que funciona y documentar su aprobación en la Chrome Web Store.
4. Crear una nueva versión más agresiva y repetir el proceso.

Esto supondría un ejemplo concreto, posible y tangible de distribución de información falsa sin que sea detectado por antivirus, firewalls o la revisión de Google, demostrando que las *fake news* tienen lugar en el Malware. Así mismo las extensiones de Chrome, una vez son aprobadas para ser publicadas y distribuidas, no son detectadas con facilidad como software indeseable. Esto quiere decir que metodológicamente debemos demostrar que puede usarse una extensión de Google Chrome para distribuir información falsa.

Finalmente decidimos acotar la noticia falsa que será distribuida. En este caso inyectamos únicamente información falsa relacionada con el Coronavirus utilizando mensajes específicos. El motivo principal para esta decisión recae sobre la necesidad de establecer puntos de control ya que no tenemos un conocimiento exacto sobre las medidas que toma Google para revisar las extensiones que enviaremos para publicar. Dado que no tenemos dicho conocimiento debemos ser metódicos con la información que subimos para demostrar el punto que queremos demostrar: que es posible inyectar noticias falsas utilizando extensiones de Google Chrome.

Todas las extensiones creadas durante esta investigación pueden ser descargadas desde Google Drive[33].

Escenario hipotético y su viabilidad

Esta investigación -como está mencionado en su introducción- se basa en un escenario hipotético para el cual se intentará probar su viabilidad en la práctica asumiendo únicamente las herramientas disponibles hoy. En este caso se plantea la existencia de una organización dedicada a la distribución de contenidos mediante técnicas inusuales como puede ser el hackeo -utilizar herramientas de forma inusual para lograr un objetivo[34]-, el spam -enviar información no deseada de forma masiva[35]-, o anuncios publicitarios en redes habituadas por sitios de baja calidad. Esta organización tiene distintos activos que le permiten tener acceso a millones de dispositivos a nivel global; en este caso por activos se hace referencia a distintos desarrollos tecnológicos como puede ser una aplicación para ver películas piratas en Android, un popular blog para descargar software pirata o incluso una

[33] https://drive.google.com/drive/folders/1Bgw0PlEiBqNrmB_hHxi_tKwP9Y5GBS2B?usp=sharing
[34] https://www.avast.com/es-es/c-hacker
[35] https://www.eset.com/es/caracteristicas/spam/

botnet -una red de dispositivos hackeados usados para distribuir *Malware*, estafas o ciberataques, Kaspersky (2022)-.

Lo cierto es que organizaciones de este tipo existen; por ejemplo, McAfee ha detectado distintas *Botnet* como es el caso de una llamada "Nitol", la cual tiene como objetivo realizar ataques DDOS, Garcia (2022). La literatura científica alrededor del tema es diversa y según la evidencia existente, la forma en la que las propias estructuras criminales se organizan varía de forma significativa.

> *(...) the majority of the structures used based on their knowledge, contacts, targets and goals. Due to the illegal genre, it is hard to know what kind of structures are the most common and effective ones, however the larger and more specific knowledge base of the group, the more money and information is stolen. For the understanding of this paper, those five categorized groups have not specified the member's goals and operations. The structure and members vary a lot, the majority of the structures mentioned in this paper are correct. [la mayoría de las estructuras utilizadas en función de sus conocimientos, contactos, objetivos y metas. Debido al género ilegal, es difícil saber qué tipo de estructuras son las más comunes y efectivas, sin embargo, cuanto mayor y más específica sea la base de conocimientos del grupo, más dinero e información se robará. Para la comprensión de este documento, esos cinco grupos categorizados no han especificado los objetivos y operaciones de los miembros. La estructura y los miembros varían mucho, la mayoría de las estructuras mencionadas en este documento son correctas.]* Jacobsson (n.d.)

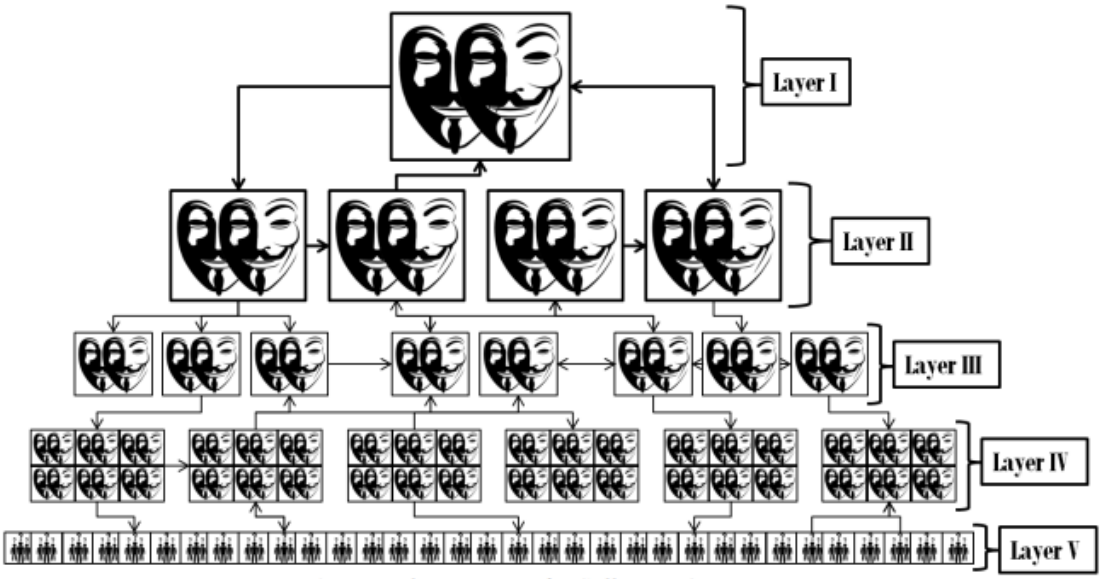
Jacobsson menciona un aspecto importante y es la capacidad que tienen individuos y organizaciones de comprar a través de la *Darknet* -sitios ocultos a los cuales no se puede acceder mediante métodos tradicionales como lo es un buscador[36]- lo que es en sí una evidencia empírica tanto de la existencia de estas organizaciones como del acceso "público" que tienen sus servicios.

[36] https://www.redseguridad.com/actualidad/cibercrimen/darknet-que-es-y-como-se-accede-a-ella_20210426.html

Además de esto Jacobsson (n.d.) menciona distintas estructuras bajo las cuales este tipo de organizaciones criminales pueden generarse, en este caso llama la atención dos tipos: las organizaciones colectivas y las organizaciones de negocio.

Figura 2

Diagrama de una organización colectiva

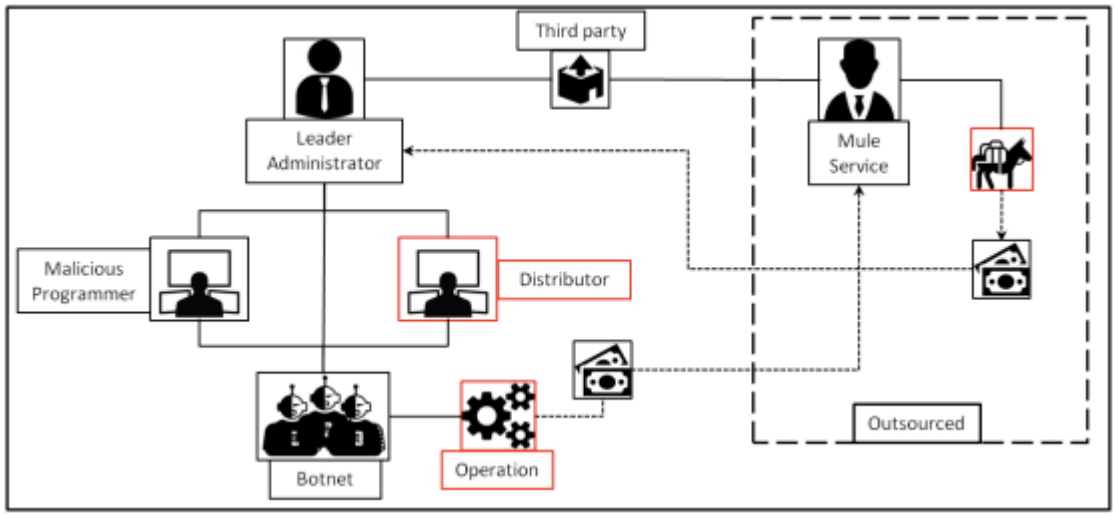


Fuente: Jacobsson (n.d.)

En cortas palabras Jacobsson sintetiza a las organizaciones colectivas como una serie de capas -5 en total- donde las responsabilidades, roles y funciones son diferentes en cada una de ellas iniciando por la capa 1 la cual es donde se concentra el liderazgo, seguido de la capa 2 que está compuesta por las personas con habilidades de ingeniería social y tecnológica, la capa 3 que incluye a los líderes regionales, la capa 4 que serían personas con habilidades informáticas que participan en algunas operaciones y terminando en la capa 5 que es la de las personas que simpatizan con la visión y objetivos de la organización.

Figura 3

Diagrama de una organización de negocio mediana



Fuente: Jacobsson (n.d.)

Tal como es posible ver en las organizaciones de negocio, la jerarquía es mucho más directa y el incentivo no es ideológico sino monetario. Son organizaciones que se dedican a la creación y renta de activos maliciosos.

Es evidente que organizaciones de este tipo existen y que tienen incentivos, sean ideológicos o económicos, suficientes para aplicar en el caso hipotético de este estudio.

Continuando con la descripción del caso, esta organización ha creado una extensión para el navegador Google Chrome con una amplia distribución en España. El éxito de esta extensión recae en que cumple una función popular entre todos los usuarios españoles: reproducir un sonido de lluvia mientras leen sus noticias favoritas. En este caso hipotético el público español ama esta extensión porque encuentran al sonido de lluvia como un factor relajante que fomenta la lectura gracias a un aumento en su concentración.

Lo que no saben es que esta extensión, además de cumplir con la función prometida de reproducir un relajante sonido de lluvia mientras leen, tiene también otras funciones ocultas completamente maliciosas siendo una de ellas la de alterar el contenido noticioso que leen en sus dispositivos. Es decir: esta extensión altera textos, imágenes, enlaces y títulos existentes en distintos periódicos españoles sin que ningún usuario sea notificado de ello.

Casos similares ya existen en la actualidad. Uno de los más populares recientemente es el de Netflix Party, una extensión que permite ver películas de Netflix con amigos a distancia ya que sincroniza el video de todos los dispositivos. Tras la pandemia de COVID-19 en el año 2020 es fácil entender por qué una extensión como esta puede ser popular, ya que con el confinamiento era imposible estar presencialmente en casa de amigos y familiares para ver una película, y "verla al mismo tiempo" es un problema, ya que no es posible saber si todos "están viendo lo mismo en pantalla". Justamente ese es un problema que esta extensión resuelve.

El caso de Netflix Party fue reportado como un claro caso de *Malware* al insertar enlaces y cookies publicitarias en los dispositivos de los usuarios con el fin de generar comisiones en sitios como Amazon, Peterson (2022). Es importante además entender que estas extensiones son numerosas y hay varios casos similares ya reportados en el pasado[37].

De hecho, existen casos registrados de extensiones de Google Chrome dedicadas exclusivamente a usuarios en regiones específicas y con funciones específicas como puede ser la banca y las operaciones financieras:

> *Researchers are warning of a remote overlay malware attack that leverages a fake Chrome browser plugin to target the accounts of banking customers in Spain.*
>
> *Grandoreiro is a type of remote overlay banking trojan, designed to help attackers overtake devices and display a full-screen overlay image when victim accesses their online banking account. In the background, meanwhile, the attacker initiates a fraudulent money transfer from the compromised account. The Grandoreiro malware, at the heart of this attack, is commonly known for exclusively targeting banking customers in Brazil – so this latest attack shows its operators expanding to victims in new countries.*
>
> *[Los investigadores advierten de un ataque de malware de superposición remota que aprovecha un falso plugin del navegador Chrome para atacar las cuentas de los clientes bancarios en España.*

[37] https://www.youtube.com/watch?v=8piNvWJHXLY

> *Grandoreiro es un tipo de troyano bancario de superposición remota, diseñado para ayudar a los atacantes a sobrepasar los dispositivos y mostrar una imagen superpuesta a pantalla completa cuando la víctima accede a su cuenta bancaria online. Mientras tanto, el atacante inicia una transferencia de dinero fraudulenta desde la cuenta comprometida. El malware Grandoreiro, que está en el centro de este ataque, es comúnmente conocido por dirigirse exclusivamente a los clientes de la banca en Brasil, por lo que este último ataque muestra que sus operadores se expanden a las víctimas en nuevos países.]* O'Donnell (2020)

Nuevamente la experiencia empírica respalda la posibilidad del caso hipotético planteado en este estudio, es decir, la serie de sucesos necesarios en los que una organización cybercriminal tenga a su disposición millones de dispositivos infectados en España con el fin de insertar información falsa en periódicos españoles es algo que es posible que pase según los estudios y casos documentados hasta la fecha. Lo que queda por saber es: ¿Google aceptará en su tienda de extensiones para Chrome una extensión que realice justamente esta función? Eso es lo que se expone a continuación.

En definitiva el uso de *Malware* es una posibilidad real en el futuro de las *fake news* y el ejercicio investigativo planteado en esta sección es plenamente válido para exponer uno de los escenarios que comunicadores e ingenieros pueden enfrentar en un futuro.

Desarrollo y publicación

A nivel de código y desarrollo existe la hipótesis -próxima a ser verificada por cada uno de los experimentos- de que Google no tendrá la capacidad de detectar las finalidades maliciosas de la extensión que es enviada para su publicación. El motivo de ello recae sobre distintas causas, la primera de ellas es que los mecanismos de detección de *malware* normalmente no se vinculan al análisis de texto, es decir, para detectar un software malintencionado o dañino las ciencias informáticas se concentran en aspectos muy diferentes al contenido en sí mismo. La otra es que el desarrollo de esta extensión no usa mecanismos avanzados o sofisticados que suelen ser utilizados en la elaboración de *Malware* para Google Chrome. Es decir: justamente la simplicidad de la extensión y el

inusual método de ataque -la modificación textual o visual de un contenido- es lo que la hace tan difícil de detectar.

Para comprender mejor lo anterior debemos empezar por entender qué es una extensión de un navegador; según Google (2022) define a las extensiones como "software programs, built on web technologies (such as HTML, CSS, and JavaScript) that enable users to customize the Chrome browsing experience." [programas de software, construidos sobre tecnologías web (como HTML, CSS y JavaScript) que permiten a los usuarios personalizar la experiencia de navegación de Chrome.]. Esto significa que una extensión es un software que de forma voluntaria un usuario añade a su navegador web -Google Chrome- para personalizar su experiencia.

Como ha sido mencionado anteriormente, hay distintas extensiones y cada una de ellas con distintas funciones y objetivos, todas ellas se encuentran en la Chrome Web Store[38] y pueden ser instaladas con un simple clic.

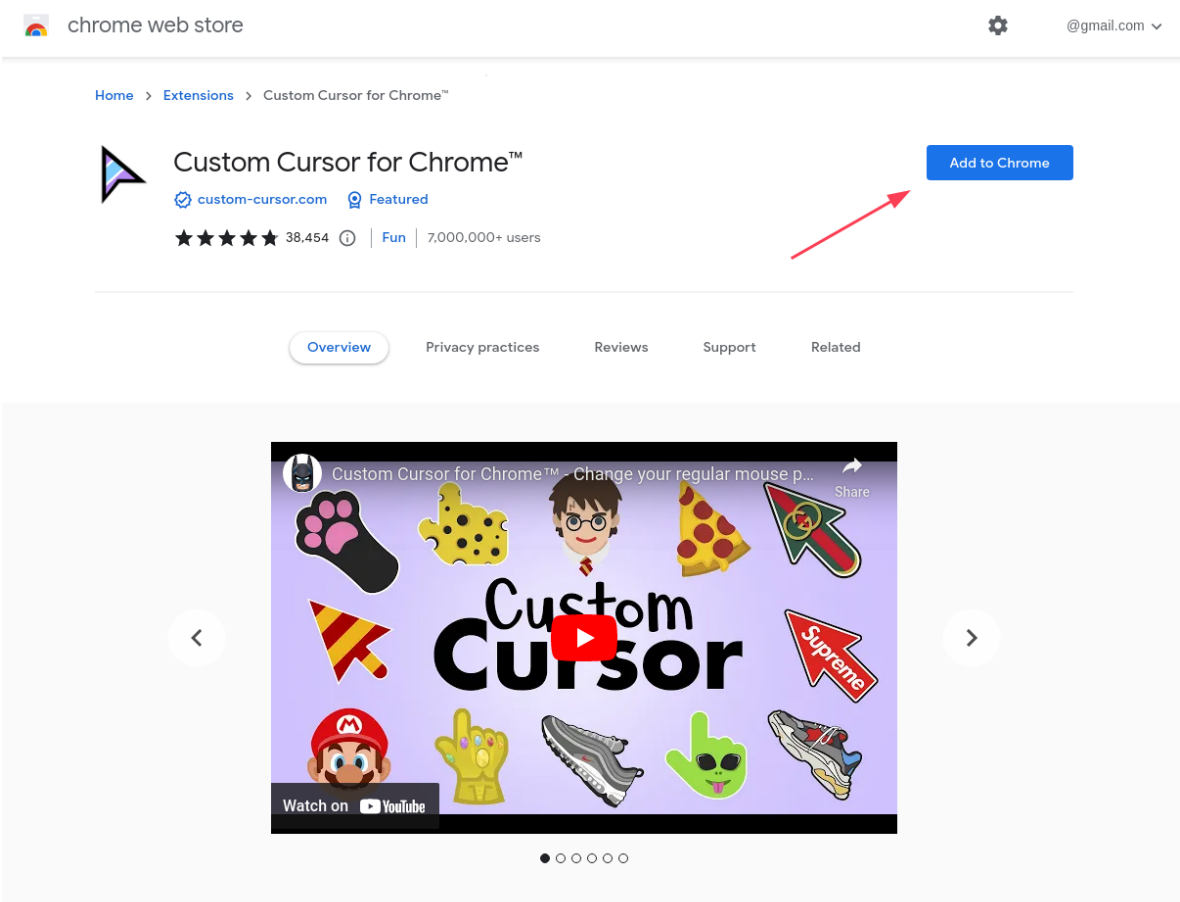
Esto hace de las extensiones algo accesible, fácil de utilizar y además usable en casi cualquier dispositivo informático -computadoras de escritorio-.

Como puede verse en la figura 4, el botón de instalación "Add to Chrome" o "Añadir a Chrome" es lo único que un usuario necesita para añadir una extensión a su navegador.

[38] https://chrome.google.com/webstore/category/extensions

Figura 4

Página de instalación de una extensión en el navegador Google Chrome



Fuente: captura de pantalla de https://chrome.google.com/webstore/detail/custom-cursor-for-chrome/ogdlpmhglpejoiomcodnpjnfgcpmgale

Lo importante de una extensión es su capacidad de utilizar determinadas funciones dentro del navegador las cuales están detalladas en el manual para desarrolladores de Google[39]. En términos generales una extensión puede acceder a la API del navegador[40] para completar sus funciones.

[39] https://developer.chrome.com/docs/extensions/mv3/intro/mv3-overview/
[40] https://developer.chrome.com/docs/extensions/reference/

Esta API tiene una enorme lista de funciones entre las cuales está el acceso a cookies, captura de pantalla, descargas, etc. No obstante, el objeto de este experimento no es detallar todas las funcionalidades existentes, sino demostrar que es posible utilizar una extensión para distribuir *fake news.*

Lo que es crucial en este proceso es que para poder publicar una extensión y hacerla disponible a millones de usuarios en todo el mundo es necesario subirla a la tienda de Google Chrome y pasar por su proceso de aprobación. Dicho proceso de aprobación es desconocido y solo Google sabe internamente cómo es efectuado. La información disponible actualmente se limita al manual de desarrolladores de la Chrome Webstore[41].

En dicho manual Google expresa distintas medidas de seguridad tomadas en cuenta al momento de hacer disponible cualquier extensión en su navegador, pero no habla directamente sobre el proceso de aprobación de cada extensión.

En todo caso, como está establecido al inicio de este capítulo, Google cada vez está más preocupado por la seguridad de su navegador y sus actualizaciones cada vez se enfocan más en detectar extensiones dañinas. Por ello en este experimento solo será utilizado un método muy sencillo: la modificación del DOM.

El DOM según Mozilla (2022) representa un documento -lo que es visible en el navegador del usuario- a través de objetos y nodos. En palabras simples, el DOM es el contenido que vemos en la web, simplemente está representado por un código que es modificable. Esto significa que si alteramos el DOM de una página web, su contenido visible será diferente, este ejemplo lo podemos visualizar en la figura 5.

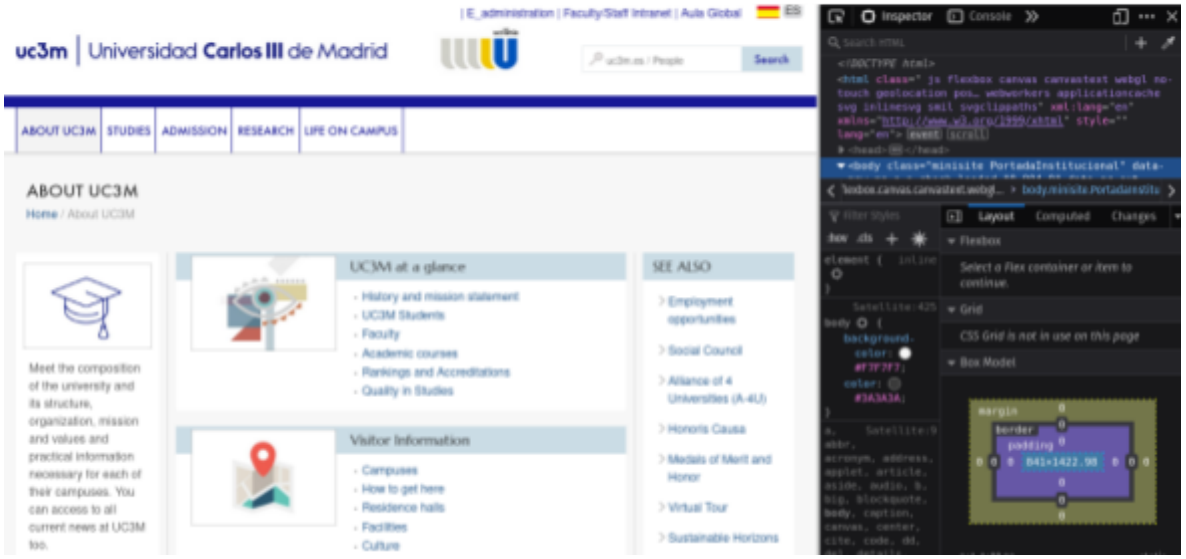
[41] https://developer.chrome.com/docs/extensions/mv3/sandboxingEval/

Figura 5

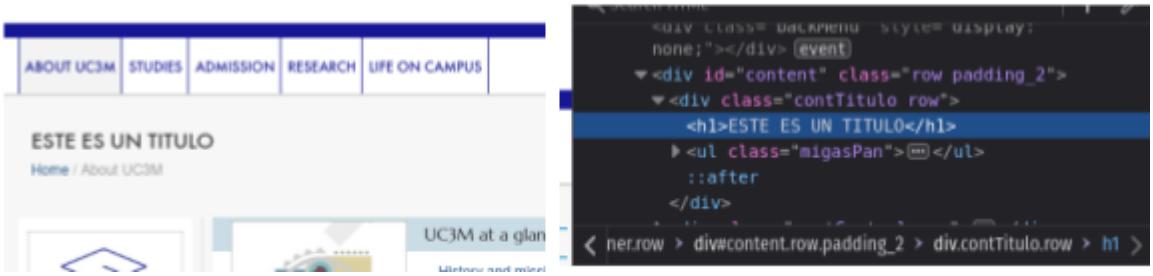
Instrucciones sobre cómo alterar el DOM en una página web

1. Puede abrirse cualquier página web

2. Al oprimir F12 es posible ver las herramientas de desarrollador



3. Al editar un contenido del DOM, este cambio será visible en el navegador web



Fuente: elaboración propia

Lo cierto es que el proceso de edición del DOM de una página web es muy sencillo, y determinadas ediciones no suelen ser tomadas en cuenta como una amenaza de seguridad como puede ser la simple edición de texto.

Por este motivo a nivel de desarrollo nos centramos en simplemente usar código dedicado a modificar el DOM de las páginas que el usuario visita y el objetivo es que precisamente Google subestime el problema de seguridad que esto puede suponer y termine por aprobar esta extensión.

Versión 0.1.5

Para el desarrollo de la primera versión fueron implementados algunos cambios bastante sencillos sobre el código original de Joaquín Vicente, algunos de ellos meramente cosméticos como los íconos de la extensión y su nombre (nuestra intención es diferenciarla al máximo posible de la extensión original) además de crear una descripción, título y resumen de lo que sería la funcionalidad de esta extensión:

- **Título**
  Sonido de lluvia personalizado
- **Resumen**
  Sonido de lluvia personalizado basado en las noticias que lees
- **Descripción**
  Con esta aplicación tendrás sonidos de lluvia personalizados basado en el contenido noticioso que estés leyendo.

Es importante considerar que la primera versión no podía incluir ningún cambio demasiado llamativo o evidente ya que cuanto más notoria sea la función de insertar contenido malicioso en la extensión, más fácilmente será para Google detectarlo y, por ende, impedir el avance del experimento. La metodología de este experimento necesita de un cambio gradual en el código para hacerlo más agresivo en futuras versiones.

Los otros cambios aplicados son visibles solo en el código fuente, el más importante de ellos es un código sencillo con una función específica: insertar el texto “(enfermedad terminal)” justo después de la palabra “Coronavirus”.

```
replaceOnDocument(/\143\157\162\157\156\141\166\151\162\165\163/g
, "\143\157\162\157\156\141\166\151\162\165\163
\50\145\156\146\145\162\155\145\144\141\144 terminal)");
```

Fuente: Elaboración propia

La función "*replaceOnDocument*" lo único que hace es buscar la palabra "Coronavirus" y reemplazarla por "Coronavirus (enfermedad terminal)" dentro de la página que el usuario esté leyendo. De forma intencional este código no inserta de forma literal el texto completo, sino que ha sido codificado en Octal -un sistema numérico de 8 dígitos[42]- con el fin de dificultar su lectura y su revisión por parte de un ser humano o de una máquina.

Por ejemplo, la palabra "coronavirus" es codificada en octal de la siguiente manera: "143 157 162 157 156 141 166 151 162 165 163". Aunque parezcan números sin sentido, para el ordenador este texto será identificado como la palabra "coronavirus". Esta función es útil ya que si es posible evitar incluir la palabra "coronavirus" será posible evitar que un ser humano que haga una posible revisión manual del código pueda saber de inmediato de qué se trata exactamente esta extensión.

Además de esto es añadido código sin función alguna dentro de la extensión solo con el fin de confundir a cualquier revisor -asumiendo que alguien podría revisar manualmente nuestro código; por tanto, es conveniente el añadir comentarios y contenido que aparentemente sea necesario para la funcionalidad principal de la extensión aunque, en realidad, no tenga influencia alguna sobre la extensión en sí o su funcionamiento:

```
//selector de música basado en contenido de lavanguardia
var myurl = "www.l\141v\141ngu\141rdi\141.com";
   var currenturl = window.location.hostname;
   if(myurl === currenturl) {
```

[42] https://www.techopedia.com/definition/6145/octal

```
      var lluvia = "sonido2.mp3"
       }
   else {
   var lluvia = "rain.mp3"
   }
      var status = 'none'
var audio = document.createElement('audio')
// Rain recording by Matt Barnard  / https://soundcloud.com/mattt
audio.setAttribute('src', lluvia)
audio.setAttribute('loop', 'true')
  function audio_control() {
    switch (status) {
      case 'none':
        audio.play()
        audio.volume = 0.5
        status = 'low'
        break
      case 'low':
        audio.volume = 1
        status = 'high'
        break
      case 'high':
        audio.pause()
        status = 'none'
        break
    }
    chrome.browserAction.setIcon({
      path: {
        "19": "images/" + status + "19px.png",
        "38": "images/" + status + "38px.png"
```

```
      }
    })
  }
chrome.browserAction.onClicked.addListener(audio_control)
```

Fuente: Elaboración propia.

Finalmente se ha determinado dentro del archivo de manifiesto un permiso explícito que solicita al navegador permisos para modificar los contenidos de lavanguardia.com

```
  "content_scripts": [{

    "js": ["content.js"],
    "matches": ["https://www.lavanguardia.com/*"]
  }]
```

Fuente: Elaboración propia.

Esta última parte es fundamental ya que es la que permite a la extensión acceder y modificar el DOM de *La Vanguardia*. Sin este permiso sería imposible el correcto funcionamiento de esta extensión.

Con esto es posible asegurarse de cumplir con tres aspectos fundamentales:

1. El código en su simplicidad se hace relativamente fácil de revisar y no incluye ninguna funcionalidad complicada que implique una sospecha por parte de cualquier revisor, lo que debe de aumentar las posibilidades de aprobación en etapas tempranas.
2. La ejecución de la extensión está limitada a un solo un dominio: lavanguardia.com, de esta forma la información falsa solo es visible en este sitio web y no en ningún otro.

3. La extensión cumple con su función: introducir información falsa dentro de un sitio web confiable.

Intentos de publicación

Tabla 1

Intentos de publicación de la versión 0.1.5

Visibilidad de publicación solicitada	Fecha de solicitud de publicación	Resultado	Fecha de publicación
Privado	2021-11-01	Publicado	2021-11-01
Deslistado	2021-12-06	Publicado	2021-12-06
Público	2021-12-06	Publicado	2021-12-06

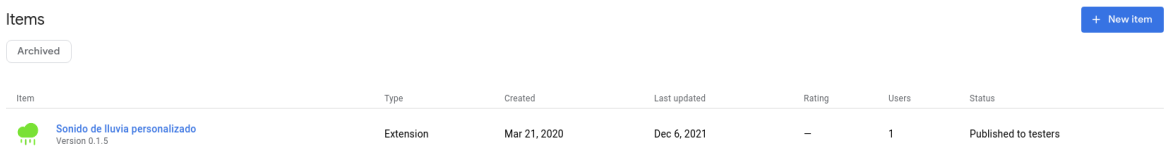
Fuente: elaboración propia

Como es posible ver en esta tabla, esta primera versión de la extensión fue aprobada en todas sus etapas, tanto en la visibilidad "privada" -limitada para que sea visible únicamente al autor de la extensión-, "Deslistada" -limitada únicamente a aquellas personas que tienen el enlace de la extensión en la tienda de Google Chrome- y en "Público" -visible para cualquier usuario que visite la tienda de extensiones de Google Chrome.

Esto significa que la primera versión publicada en la tienda de Google fue aprobada plenamente para su uso en todo internet.

Figura 6

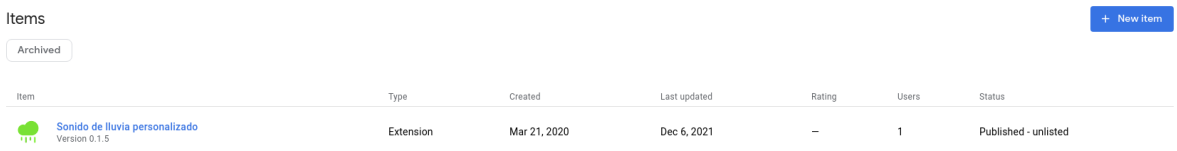
Extensión aprobada para testers en su versión 0.1.5



Fuente: elaboración propia

Figura 7

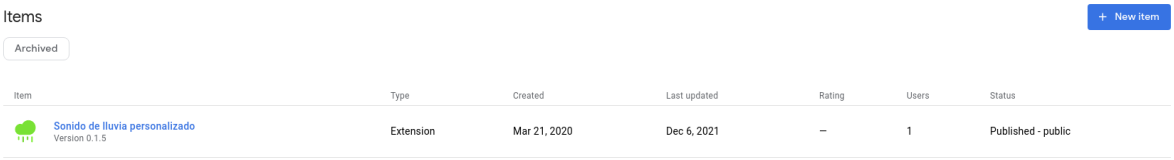
Extensión aprobada de forma privada en su versión 0.1.5



Fuente: elaboración propia

Figura 8

Extensión aprobada de forma pública en su versión 0.1.5



Fuente: elaboración propia

La aplicación en su versión 0.1.5 fue aceptada y publicada por Google. En un simple video está expuesto el cómo funciona, además de mostrar la simple instalación en un navegador de Google Chrome[43]. Con este video es mucho más fácil entender cómo funciona la extensión y cuál es su verdadero impacto en el dispositivo de un usuario.

[43] https://www.youtube.com/watch?v=pyeroJ-DCl8

Figura 9

Visualización del funcionamiento de la extensión de Google Chrome en su versión 0.1.5

Fuente: elaboración propia.

Versión 0.1.9

Logrado demostrar que era posible insertar contenido modificado dentro de lavanguardia.com con la versión 0.1.5 en esta nueva versión la extensión es más compleja y añade funciones de insertar enlaces externos que expanden y profundizan la información falsa. La versión 0.1.9 tiene como función principal insertar un enlace con información falsa tal como está explicado en video[44].

Esta nueva versión tiene varias novedades, la primera de ellas es la actualización del archivo manifest en donde pasamos a la versión 3[45] -esto ayuda a demostrar la viabilidad de estas aplicaciones malintencionadas aún con las últimas actualizaciones de Google-.

[44] https://www.youtube.com/watch?v=04MR_PKSY34
[45] https://developer.chrome.com/docs/extensions/mv3/intro/

```
{
  "manifest_version": 3,
  "name": "Sonido de lluvia personalizado",
  "version": "0.1.9",
  "description": "Sonido de lluvia personalizado basado en las
noticias que lees",
  "background": {
    "service_worker": "rain.js"
  },
  "content_scripts": [{
    "js": ["content.js"],
    "matches": ["https://www.lavanguardia.com/*"]
  }]
,
"host_permissions": [
  "https://www.lavanguardia.com/",
  "*://*/*"
],
  "icons": {
    "128": "images/icon-128px.png",
    "48": "images/icon-48px.png"
  },
  "browser_action": {
    "default_icon": {
      "19": "images/none19px.png",
      "38": "images/none38px.png"
    }}}
```

Fuente: Elaboración propia. Archivo manifest.json de la versión 0.1.9

Además de esto se han modificado los archivos de la extensión para insertar con un código ofuscado en octal un texto modificado que dice “vacunas para el Coronavirus tienen riesgos para la salud mental” el cual enlaza a una publicación con noticias falsas en Facebook.

```
const replaceOnDocument = (pattern, string, {target =
document.body} = {}) => {
    target.querySelectorAll("p, h1, h2, h3, h4").forEach(element
=> {
        element.innerHTML =
element.innerHTML.replace(/C\157\162\157\156\141\166\151\162\165\
163/g, '
\74\141\40\150\162\145\146\75\42\150\164\164\160\163\72\57\57\167
\167\167\56\146\141\143\145\142\157\157\153\56\143\157\155\57\160
\145\162\155\141\154\151\156\153\56\160\150\160\77\163\164\157\16
2\171\137\146\142\151\144\75\61\63\60\60\61\70\65\60\66\61\62\66\
62\70\60\46\151\144\75\61\60\60\60\67\63\65\64\63\71\64\66\66\63\
62\42\76\166\141\143\165\156\141\163 para el
C\157\162\157\156\141\166\151\162\165\163
\164\151\145\156\145\156\40\162\151\145\163\147\157\163\40\160\14
1\162\141\40\154\141\40\163\141\154\165\144\40\155\145\156\164\14
1\154</a> ')
    })
};
```

Fuente: Elaboración propia.

Estos cambios en el código fuente son sencillos e introducen algunos aspectos que hacen de la modificación de contenido algo más agresivo. En primer lugar, es añadida la función “querySelectorAll” para los elementos de párrafo y encabezados -del 1 al 4- con el fin de insertar el contenido incluso en elementos llamativos de la página como puede ser el título principal de una noticia. Además de ello se deja de usar la función “replaceOnDocument” y en su lugar se pasa a usar “element.innerHTML.replace”. Este cambio es importante ya que “replaceOnDocument” solo permite cambiar elementos textuales de la página, mientras que

“innerHTML.replace” permite cambiar elementos HTML tales como hipervínculos lo cual es, precisamente, lo que empieza a hacer esta nueva extensión ya que ahora introduce un enlace hacia una noticia falsa externa a *lavanguardia.com*.

Esta nueva versión se hace notoriamente más agresiva que la anterior al insertar un enlace. De hecho, en algún momento se consideró que esta característica haría de la extensión un recurso que sería detectado y bloqueado automáticamente por Google ya que el insertar un enlace no autorizado en una página web puede suponer precisamente un problema de seguridad severo. Por ejemplo, si una extensión insertara un enlace malicioso en una página financiera como la de un banco podría ser un desastre.

De forma sorpresiva, y como será observado en los resultados, esta versión de la extensión también fue aprobada por Google pese a ser un desarrollo considerablemente más agresivo.

Es posible descargar el código fuente para verificar el comportamiento de esta extensión[46]. Finalmente, tras publicar esta nueva versión verificamos que Google la aprueba y la hace pública en cuestión de minutos.

[46] https://drive.google.com/file/d/1yO64IElppwAykQIGLB1H6KsIvr1GreI5/view?usp=sharing

Figura 10

Página a la que enlaza la versión 0.1.9 de la extensión

Fuente: elaboración propia

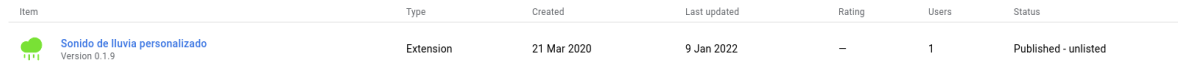
Figura 11

Extensión publicada para testers en su versión 0.1.9

Fuente: elaboración propia

Figura 12

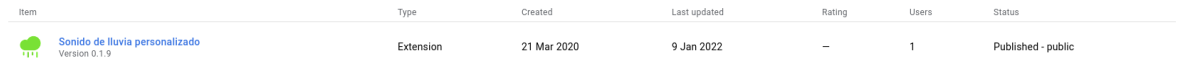
Extensión publicada de forma privada en su versión 0.1.9



Fuente: elaboración propia

Figura 13

Extensión publicada de forma pública en su versión 0.1.9



Fuente: elaboración propia

Tabla 2

Intentos de publicación de la versión 0.1.9

Visibilidad de publicación solicitada	**Fecha de solicitud de publicación**	**Resultado**	**Fecha de publicación**
Privado	2021-01-09	Publicado	2021-01-09
Deslistado	2021-01-09	Publicado	2021-01-09
Público	2021-01-09	Publicado	2021-01-09

Fuente: elaboración propia

Figura 14

Visualización del funcionamiento de la extensión de Google Chrome en su versión 0.1.9 en donde una noticia es manipulada sin conocimiento del lector

Fuente: elaboración propia

Es difícil determinar por qué Google aprobó esta versión considerando que la sola existencia de un enlace externo insertado por una extensión debería ser un aspecto a revisar. Aquí solo existen dos posibilidades: una es que Google efectivamente detectó que esta extensión inserta un enlace en la web, pero sus sistemas no lo evaluaron como un elemento dañino ya que la página de destino es Facebook, o bien Google nunca detectó que el código de esta extensión inserta un enlace en el DOM de la página de *La Vanguardia*. Sea cual sea la causa es simplemente una realidad que suscita preocupación.

Versión 0.1.11

Tras observar en las versiones anteriores la baja resistencia de la Chrome Web Store en el proceso de revisión y aprobación se decidió dar un salto mucho más agresivo en la versión

0.1.11 al inyectar directamente una noticia falsa en una URL inexistente de lavanguardia.com. Para lograr esto se ha inventado una URL inexistente que denominamos como "https://www.lavanguardia.com/covid-alerta-mundial", cualquier usuario que intente acceder a esta URL verá que se trata de una página sin contenido -un error 404- no obstante en nuestra aplicación insertamos un código encargado de insertar contenido en esta página y hacer parecer que realmente se trata de una página real.

```
const replaceContentFakeURL = () => {
      if (window.location.href ===
'https://www.lavanguardia.com/covid-alerta-mundial') {
      document.title = "Las vacunas contra el Coronavirus causan
cáncer";
      return document.body.innerHTML = `<html lang="es"><head>
      <meta http-equiv="Content-Type" content="text/html;
charset=utf-8">
```

Fuente: elaboración propia.

El código expuesto hace una validación muy sencilla en la que revisa la URL en la que se encuentra el usuario. En caso de ser la URL deseada reemplazará su contenido y título. Para esta versión no se aplicó una ofuscación en ninguna parte del texto y se dejó en texto plano todo el contenido de la noticia con el fin de poner a prueba el sistema de verificación de extensiones de Google y verificar con mayor claridad si existe una revisión asociada al texto incluído dentro de las extensiones. Pese a ser una versión con funciones más explícitas y llamativas que las anteriores la Chrome Web Store la aprobó sin ningún tipo de resistencia.

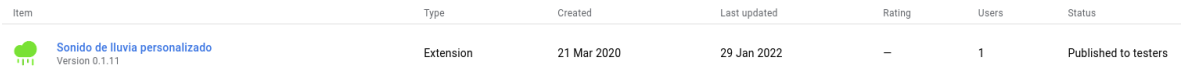
Figura 15

Visualización del funcionamiento de la extensión de Google Chrome en su versión 0.1.11 en donde una noticia falsa es insertada en La Vanguardia sin conocimiento del lector

Fuente: Imagen de elaboración propia.

Figura 16

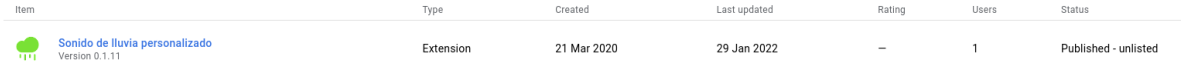
Extensión publicada para testers en su versión 0.1.11



Fuente: elaboración propia

Figura 17

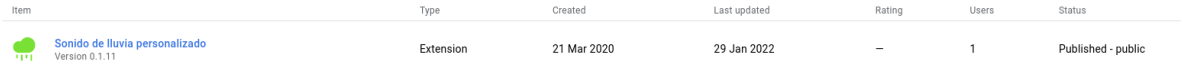
Extensión publicada de forma privada en su versión 0.1.11



Fuente: elaboración propia

Figura 18

Extensión publicada de forma pública en su versión 0.1.11



Fuente: elaboración propia

Intentos de publicación

Visibilidad de publicación solicitada	Fecha de solicitud de publicación	Resultado	Fecha de publicación
Privado	2021-01-29	Publicado	2021-01-29
Deslistado	2021-01-29	Publicado	2021-01-29
Público	2021-01-29	Publicado	2021-01-29

Por qué Google aprobó todas las extensiones desarrolladas

No termina de ser una sorpresa el hecho de que Google no detectara en ningún momento ninguna de las versiones como un elemento sospechoso o, como mínimo, como un objeto a ser revisado con mayor exactitud. Cada una de las versiones de esta extensión insertaba la información falsa de distintas maneras: alterando el texto, alterando el HTML o insertando contenido completamente nuevo.

Google no es específico con sus métodos de revisión, pero la explicación más simple con lo obtenido en este experimento es que sus métodos de detección no se enfocan en detectar o evaluar a profundidad extensiones cuya única función es insertar contenido en el DOM de una página web.

Es cierto que no existen casos en donde se hayan usado estos métodos a escala con el fin de cometer ciberdelitos y esto sea lo que explica la nula resistencia que fue encontrada al publicar estas extensiones de Google Chrome.

Probablemente Google solo tenga como enfoque principal revisar y detectar extensiones que aplican ataques o acciones malintencionadas ya conocidas y que precisamente lo hecho en este ejercicio, al ser una actividad no vista antes, no pase de momento por ninguno de los filtros del gigante tecnológico.

Conclusiones

El experimento planteado en esta investigación trae resultados interesantes que a su vez demuestran que la distribución de contenido desinformativo a través del navegador de Google Chrome y su tienda de extensiones es perfectamente posible. Dado que nuestro objetivo no es exponer una vulnerabilidad informática sino demostrar que es posible distribuir *fake news* haciendo uso de software malicioso pudimos ver a través de distintas versiones -que fueron cada vez más agresivas- que todas las variaciones fueron revisadas y publicadas en la Chrome Web Store y estuvieron disponibles para cualquier usuario a nivel global.

Dado que no existe una documentación pública exacta sobre los métodos de detección de malware de Google para la Chrome Web Store no podemos decir con exactitud bajo qué

escrutinio técnico la extensión de este estudio y cualquiera de sus versiones fue revisada. Lo que sí evidenciamos es una carencia absoluta de revisión de textos y contenidos; es decir, el sistema de revisión de Google no hace una lectura -ni directa ni indirecta- del contenido textual de la extensión. Esto deja de manifiesto -al menos en el contexto exacto de nuestro estudio- que no hay mecanismo alguno para detectar contenido desinformativo en una pieza de software, demostrando así que el escenario hipotético planteado inicialmente es factible: utilizar software malicioso para distribuir contenido desinformativo o *fake news* es posible.

Esto expone un peligro latente para el cual muchas tecnologías actuales no están preparadas, para empezar los navegadores web a pesar de contar con sistemas de encriptación y certificación ssl no incluyen ningún tipo de protección adicional que permita certificar que el contenido que el usuario consume en su dispositivo es el mismo que el servidor tiene publicado. En otras palabras: no hay mecanismos de protección para el contenido consumido por el usuario de manipulación externa.

Futuras investigaciones

Lo expuesto en este artículo permite prever lo que podrían ser futuras investigaciones y desarrollos a nivel informático en relación a las *fake news*. En primera instancia el análisis de contenido debería ser una futura extensión dentro del conocimiento correspondiente a la seguridad informática ya que en distintos escenarios el software malicioso podría estar diseñado para inyectar o modificar contenido. Si bien no es posible saber "qué contenido es falso" sí podrían encontrarse métodos de detección de software diseñado para alterar el mismo.

También debería existir un seguimiento sobre la evolución del malware alrededor del contenido indeseado como la publicidad y su posible relación con las noticias falsas. Como ha sido expuesto anteriormente la relación entre malware y publicidad o contenido indeseado no es algo nuevo lo que hace posible que los métodos actuales de distribución de adware evolucionen directa o indirectamente hacia la distribución de contenido desinformativo.

CONCLUSIONES FINALES Y PRÓXIMAS LÍNEAS INVESTIGATIVAS

Este estudio investigativo supone una exploración demasiado extensa y profunda sobre la distribución de las *fake news* a través de los medios digitales en donde fue posible evidenciar cómo es la misma distribución de estos mensajes el principal factor de éxito para este tipo de mensajes.

La distribución del mensaje tiene un estrecho vínculo con la tecnología y sólo entendiendo la tecnología y su funcionamiento es posible lograr una distribución efectiva del mensaje, o también impedir la distribución del contenido desinformativo. Saber cómo funcionan los algoritmos es el mejor camino para impedir la proliferación de la desinformación, al menos en la era digital que se vive al momento de escribir esta tesis doctoral.

Todo tiene un espectro de complejidad importante al involucrar la tecnología ya que cualquier ámbito tecnológico tiene detrás de sí una serie de aspectos que suelen ser ignorados pero que son importantes para explicar cualquier fenómeno o comportamiento referente a la distribución de un mensaje -por ejemplo, entender el funcionamiento del *Machine Learning* para comprender el comportamiento del *News Feed* de Facebook en lo que a detección de información falsa se refiere-.

Esta complejidad es lo que abre un abanico investigativo que hasta hoy tiene una fuerte concentración en la detección de la *Fake New* lo que trae una oportunidad de investigar en nuevas áreas como puede ser precisamente el software como método de distribución.

Antes de continuar es importante retomar las tres principales preguntas planteadas en la introducción de este trabajo investigativo:

- ¿Cómo logran distribuir las noticias falsas a través de Google y Meta?
- ¿Existen cambios importantes tras las elecciones presidenciales de 2016 en los Estados Unidos que afectasen la distribución de desinformación?
- ¿Es posible utilizar software malicioso para la distribución de *fake news*?

Como se ha visto a lo largo del trabajo, existe un constante esfuerzo por parte de los autores de las *fake news* para conseguir una distribución de su mensaje hacia la mayor audiencia

posible y el principal método para lograr esto es a través de distintas técnicas que cambian según el medio bajo el cual quieren lograr dicha distribución.

En el caso de Facebook es evidente que la conclusión general es que la distribución de noticias falsas se logra a través de dos grandes mecanismos; El primero es el de "aceptar y usar los formatos y contenidos que Facebook -o cualquier otra red social de Meta- más favorece"; claramente la visibilidad de ciertas publicaciones depende en gran medida de lo que el algoritmo de Facebook quiere favorecer -o castigar-, como fue encontrado en el caso de los periódicos españoles el éxito de la gran mayoría de las publicaciones depende de su formato -el video- y esto es algo que los distribuidores de noticias falsas saben a la perfección.

Esta primera técnica puede sintetizarse en uno de los términos que ya fueron adoptados por los profesionales de la comunicación y el márketing digital: NFO -News Feed Optimization-, esto es porque el único objetivo de esta práctica es publicar contenido con el fin de obtener el mayor alcance posible, y esto se logra precisamente "entendiendo qué quiere el algoritmo de Facebook"; es decir, no es posible obtener un gran alcance si el algoritmo sencillamente no desea priorizar un contenido particular en el *News Feed.*

Para los autores de las *fake news* es por tanto importante entender el estado actual de los algoritmos y estar al tanto de las últimas tendencias en tópicos, formatos y contenidos populares entre la población con el único fin de obtener precisamente una distribución amplia y generalizada en Facebook.

Tal como distintos casos evidencian esto trae siempre nuevas técnicas de distribución para las cuales los autores de las *fake news* se adaptan de forma constante, es una auténtica carrera para impedir que Facebook frene sus avances.

El segundo mecanismo es el de uso de anuncios, grupos y herramientas de distribución de contenidos que Facebook pone a disposición de cualquier anunciante o empresa. Este mecanismo estuvo presente en el caso de las elecciones presidenciales de los Estados Unidos del año 2016, etapa en la que los anuncios de Facebook fueron clave para la distribución de *fake news.* Si no usan las herramientas propias de Meta distribuir cualquier

contenido se haría imposible -por ejemplo, es más fácil distribuir una *Fake New* usando el formato de video de Facebook que un enlace externo-.

Otro aspecto muy necesario está en el "disfrazar" u "ocultar" el contenido desinformativo. Tal como se ha visto tanto en las investigaciones académicas vigentes como en el comportamiento histórico de Facebook existe un importante interés en la detección de contenido engañoso, desinformativo o incluso de baja calidad -contenido de clickbait, que promueve las interacciones, que es polémico, etc-, un formato de contenido que en efecto es usado por distintas fuentes de *fake news* ya que evidentemente la viralización de estos contenidos es una de sus fuentes de ingresos. Esto hace que los autores de las noticias falsas deban trabajar constantemente en transformar su forma de redactar y publicar sus contenidos, todo con el fin de que Facebook no los detecte. En pocas palabras: la propia transformación del contenido es clave para el futuro del mismo dentro de Facebook.

El caso de Google es completamente diferente, ya que al tratarse de un buscador la forma en la que permite la distribución de las *fake news* es de una forma distinta a la de Meta. Mientras que Meta es una fuente "única" en donde todos publican y Facebook decide qué publicación es visible y recomendada, Google es una fuente "distribuida" ya que cada usuario decide qué quiere buscar -con esto no se quiere decir que Google no sea un monopolio o que su control en los resultados de búsqueda no sea centralizado-.

Estos retos técnicos han hecho que tanto Google como Meta tengan incluso acercamientos diferentes frente al problema de las *fake news,* mientras Meta cuenta con toda la información centralizada y bajo su control -todas las publicaciones ocurren dentro de Facebook- Google debe lidiar con todo tipo de formatos, fuentes y tecnologías web -Google debe rastrear y *parsear* toda la web-.

Esto explica el hecho de que Meta reaccione a las *fake news* con actualizaciones en sus algoritmos y con revisiones tanto manuales como basadas en *Machine Learning*. Después de todo Facebook tiene el control del formato y el modelo de datos de su propio ecosistema. Google, por otro lado, depende mucho más de terceros: no necesita "entender mejor la información de su propio sistema", más bien entender qué hay en sistemas externos al suyo. Ello lo ha llevado a trabajar en múltiples proyectos y tecnologías al

servicio de terceros -siendo el más destacado *The Trust Project*- que al largo plazo deberían facilitar precisamente el trabajo de Google al momento de detectar información falsa.

Por lo anterior la forma en la que una noticia falsa es distribuída en el buscador de Google tiene otro tipo de retos -y probablemente mucho más complejos- que en Facebook. Si en Facebook una publicación es detectada como información falsa siempre se podrá publicar una nueva, y así mismo si un grupo o página es castigado por distribuir información falsa, siempre se puede crear uno nuevo.

Es Google quien decide qué sitios web deben de ser posicionados en determinadas búsquedas. Esto tiene una naturaleza mucho más complicada ya que la posición en Google es un ciclo mucho más lento que en Facebook -mientras en Facebook lograr la viralidad de a través de un contenido es algo que podría tardar horas, en Google el posicionamiento de una página o dominio puede tomar años-. Esto hace que para lograr posicionar y distribuir noticias falsas a través del buscador de Google sea necesario construir a lo largo del tiempo portales noticiosos que pasen desapercibidos para Google.

En la comparación de distribución de contenidos entre Google y Meta queda claro que la eficacia de la distribución de cualquier contenido cambia de forma drástica dado que la forma en la que cada tecnología funciona es intrínsecamente distinta: mientras Facebook tiene una estructura centralizada en donde cada usuario consume el contenido que el *News Feed* le facilita, en Google el usuario encuentra aquello que busca. Por otro lado, en Meta hay una amplia diversidad de formatos como video, imágenes, texto, etc. Mientras que en Google todo recae sobre el posicionamiento que un recurso web recibe en determinadas búsquedas.

Esto hace que las técnicas usadas por parte de los autores de las *fake news* sea distinta según cada plataforma -algo obvio para cualquier usuario- pero hace que los mecanismos de detección y prevención de la desinformación sean muy distintos en cada plataforma, por ejemplo: el *clickbait* funciona muy bien en Facebook mientras que no tiene efecto alguno en Google.

Lo cierto es que, aunque estas formas de publicación y distribución son conclusiones extraídas de los resultados de este estudio, está claro que antes de las elecciones presidenciales del año 2016 la realidad era muy diferente.

Antes de dichas elecciones publicar y distribuir *fake news* en Facebook y en Google era una tarea mucho más sencilla por el simple hecho de que no existían controles exhaustivos de quién publica y patrocina determinada información. Es cierto que podría ser debatido e incluso investigado en el futuro si "el escrutinio y controles actuales de Google y Meta son suficientes para frenar la distribución de *fake news*", pero más allá de si los esfuerzos actuales son suficientes, es un hecho que tras la elección de Trump como presidente de los Estados Unidos dichos esfuerzos se incrementaron de forma notoria.

Claramente antes de las elecciones era mucho más sencillo lograr la distribución de un contenido desinformativo. Esto hace que a partir del año 2016 se entrase en una "carrera de información y distribución" en donde Google y Meta hacen todo lo posible por mitigar la distribución de *fake news* mientras que los autores y las organizaciones responsables de este tipo de contenidos intentan "engañar" al sistema con el objetivo de continuar obteniendo visitas y visibilidad, lo cual es el aspecto más importante para su negocio y rentabilidad.

La mejor evidencia de esto recae sobre la inversión tecnológica de Google y Meta para detectar las noticias falsas, en el caso de Meta para impedir su distribución en sus redes sociales y en el caso de Google para poder encontrar aquellos sitios y autores responsables para impedir su posicionamiento web.

El caso es que es posible ver que Meta logra materializar de una forma más efectiva distintas técnicas de detección de *fake news* en comparación a Google.

Vale la pena mencionar que las *fake news* en el ámbito digital requieren de un amplio espectro de datos para que la comunidad científica pueda llegar a resultados más extensos y complejos de los que tenemos hoy, volumen de datos que necesita ser nutrido a partir de distintas fuentes que por desgracia en la actualidad no son disponibles.

Por ejemplo, como fue observado en el marco teórico y estado de la cuestión, la gran mayoría de investigaciones basadas en la detección de noticias falsas se basan en datos ya existentes y previamente catalogados de noticias falsas, sistemas de datos que tienen varios

problemas: el primero de ellos es su rápida desactualización ya que cada día salen nuevas *fake news* con nuevas técnicas de distribución y redacción haciendo de los modelos de *Machine Learning* un caso de estudio que, aunque sea interesante, es obsoleto en el momento que se crea. Además de ello se tiene el problema del sesgo de quienes catalogan determinadas noticias en dichas bases de datos: no es un secreto que el determinar "qué es cierto y qué no lo es" no es una tarea sencilla y los sesgos cognitivos de los responsables de dicha categorización puede ser justamente un elemento que hace de los datos una fuente menos fiable de lo que debería ser, aunque esto último no es posible determinarlo.

Pero no está solo en los datos de aquello que puede ser considerado como una noticia falsa, también está en los datos de audiencia, algorítmicos y de distribución que las grandes tecnológicas tienen. Ni Google ni Meta comparten al público datos específicos sobre cómo reaccionan sus usuarios a las *fake news,* ni tampoco tienen indicadores públicos sobre cómo sus medidas tienen algún nivel de eficacia. Esto hace que el combate contra la desinformación, al menos en los servicios de las grandes tecnológicas, sea una caja opaca imposible de interpretar.

En varios capítulos de esta tesis doctoral fue imposible obtener datos concretos de Google o de Meta haciendo que muchos aspectos dependen de la interpretación de datos externos o del análisis de fuentes externas en las que sí es posible obtener información concreta. En efecto esto refuerza la idea de que el éxito o el fracaso de una *Fake New* siga dependiendo en gran medida de las acciones que una compañía privada toma. No es posible para una institución pública o para la academia aportar estudios o tomas de decisiones sustanciales si la información está salvaguardada tras una caja fuerte.

No es para poco, ningún estudio citado tiene información concreta sobre "cuántas personas vieron *fake news* en Facebook" o "cuántas noticias falsas fueron distribuidas a través de Google". Sin información ésta y cualquier otra investigación se encontrará con las mismas limitaciones. Desafortunadamente la academia debe seguir avanzando en esta situación.

Una observación interesante durante el desarrollo de esta tesis es que fue precisamente después de las elecciones presidenciales estadounidenses de 2016, cuando fue notorio un gran esfuerzo por parte de Google y Meta para combatir la desinformación. Es llamativo porque la desinformación en este punto de la historia no era un fenómeno nuevo, los

contenidos desinformativos ya eran populares en Internet, pero el simple hecho de que estos contenidos lograran tener una determinada influencia en la política norteamericana fue el factor decisivo para generar este cambio.

Es imposible saber si antes de las elecciones presidenciales Google y Meta estaban al tanto de lo que ocurría en sus propias plataformas en lo que se refiere a *fake news* lo que solo deja sobre la mesa dos posibilidades: si no sabían lo que estaba ocurriendo con la información falsa en sus propios sistemas entonces se evidencia su propia incapacidad de controlar y gestionar la masiva cantidad de contenidos que fluye a través de ellos. Por el contrario, si lo sabían, y no hicieron nada al respecto antes de las elecciones, entonces sus intenciones no están del lado de los ciudadanos sino de sus propios beneficios publicitarios.

Finalmente queda el lado tecnológico de esta investigación; ¿es posible distribuir *fake news* haciendo uso de software malicioso? La respuesta a esta pregunta debe abordarse desde varios puntos de vista ya que la misma tecnología es partícipe en distintos puntos de la cadena de información. En primera instancia la tecnología claramente puede usarse para generar información falsa a escala: es posible exponer continuamente ejemplo tras otro. Por ejemplo, si alguien quisiera generar una falsa narrativa sobre la guerra entre Rusia y Ucrania -evento en desarrollo durante la redacción de este estudio- solo sería necesario pedir a una inteligencia artificial que genere una imagen de “moscu terrorist attack” y dará como resultado la figura 1.

No puede quedar duda que la Inteligencia artificial va a tener un rol primordial en el futuro de la desinformación y que será cada vez más accesible para cualquier persona el poder crear imágenes cada vez más complejas con muchísimo menor esfuerzo.

Aquí fue posible crear dos fotografías completamente falsas que apuntan a dos noticias falsas muy sensibles en el entorno político y social de la actualidad, y fue completamente fácil y accesible con un ordenador y sin mayores conocimientos de ingeniería informática.

Figura 1

Fotografía falsa de los sucesos de un ataque terrorista en Moscú generada con Inteligencia Artificial

Fuente: elaboración propia con Stable Diffusion.

Sin lugar a dudas, los autores de las noticias falsas tienen nuevas capacidades para generar imágenes y textos que ayuden a tener una narrativa elaborada y rentable.

En lo que respecta a la elaboración de contenido basándose en Inteligencia Artificial solo es posible esperar que estos mecanismos serán más complejos, precisos y accesibles con el paso del tiempo. Es decir, estará en la capacidad de todos los seres humanos la posibilidad de generar con mayor frecuencia información y contenido con un nivel de complejidad que antes estaba únicamente al alcance de grandes compañías. Esto hace de la tecnología un pilar fundamental en el futuro de la generación del contenido desinformativo.

Por otra parte, tenemos la tecnología como facilitador de la distribución de la información en donde el ejemplo por excelencia para esta investigación fue el *Newsfeed* de Facebook. Claramente las nuevas tecnologías y servicios facilitan el acceso de cualquier persona a una enorme audiencia. Esto era algo que no era posible antes ya que toda distribución en masa

de cualquier mensaje era algo que dependía completamente de los grandes medios de comunicación -especialmente los periódicos, la radio y la televisión-. Naturalmente al existir sitios que centralizan gigantescas audiencias en donde un algoritmo es el único que decide quién ve qué contenido se hace crucial entender el comportamiento del mismo para lograr la distribución exitosa de un mensaje.

Esto no se limita únicamente al "algoritmo del *Newsfeed*". En efecto entender los algoritmos de cualquier medio digital es crucial, incluyendo sin lugar a dudas a Google y a TikTok.

Esto hace que el rol distributivo que tiene la tecnología sea un punto de cambio y de dinámicas cambiantes de cara al futuro, en donde los autores de las *fake news* harán todo lo que esté a su alcance para que el algoritmo los favorezca, mientras que las grandes tecnológicas intentarán día a tras día detectar dicho contenido para impedir su distribución.

Esto nos lleva al último rol que la tecnología tiene en todo este estudio y es la posibilidad de usar *Malware* como recurso distributivo del mensaje. Es importante entender que este método es completamente diferente al de publicar un contenido o una web, en este caso se estaría utilizando un software con funciones no deseadas con el objetivo de engañar y distribuir mensajes maliciosos.

En este caso y, sin lugar a dudas, queda claro que el *Malware* es un recurso viable para la distribución de las *fake news*. En el caso hipotético planteado en esta investigación se busca encontrar la posibilidad de distribuir una noticia falsa usando extensiones de Google Chrome, caso hipotético validado tanto por hechos históricos como por la misma prueba ejecutada durante esta investigación: la extensión que inyecta información falsa directamente en lavanguardia.com fue aprobada en todas las ocasiones por Google sin que existiera advertencia o dificultad alguna en el proceso.

Es cierto que es imposible determinar por qué esta extensión no fue detectada por Google como un *Malware* -esto solo lo pueden saber los ingenieros de Google- pero el escenario más plausible es bastante simple: no existen métodos de detección de *Malware* basados en la modificación textual de un contenido.

Si hemos visto que el *Malware* se usa para ganar dinero o distribuir spam, es perfectamente posible que éste sea usado para distribuir información falsa. En este sentido en el futuro es plausible encontrar organizaciones criminales enfocadas en este tipo de métodos para distribuir material desinformativo en la población.

Las futuras líneas investigativas posibles tras esta investigación deben estar enfocadas precisamente en la distribución del contenido desinformativo no solo en medios digitales tradicionales como Meta o Google, sino también en medios "oscuros" -o no traceables- como WhatsApp o Telegram. Así mismo profundizar en nuevos sitios como TikTok ayudarán a complementar la foto general que se tiene de los algoritmos de distribución de contenidos.

Por otro lado, está claro que es necesaria una profundización en la relación entre tecnología y *fake news* y que, además, será una relación con un alto nivel de evolución y desarrollo en los próximos años. Entender al *Malware* como parte de la cotidianidad y como un posible foco de distribución futuro ayudará a periodistas e ingenieros a encontrar más y mejores soluciones que mitiguen la distribución de contenido desinformativo.

BIBLIOGRAFÍA

Abad, C. S. (2019). La primera *Fake News* de la historia. Historia y Comunicación Social, 24(2), 411–431. https://doi.org/10.5209/HICS.66268

Adweek. (2007). *Inside Facebook, NFO (News Feed Optimization) is the new SEO – Adweek*. https://www.adweek.com/digital/inside-facebook-nfo-is-the-new-seo/

Adweek. (2007). *Facebook's News Feed Knows What You Did Last Summer – Adweek*. https://www.adweek.com/digital/facebook%e2%80%99s-news-feed-knows-what-you-did-last-summer/

Ahmed, W., Downing, J., Tuters, M., & Knight, P. (2020). Four experts investigate how the 5G coronavirus conspiracy theory began. https://theconversation.com/four-experts-investigate-how-the-5g-coronavirus-conspiracy-theory-began-139137

Alfonso, I. B.; Galera, C. G.; y Calvo, S. T. (2019). "The impact of *Fake News* on Social Science Research. Systematized bibliographic review". Historia y Comunicación Social, 24(2), 449–469. https://doi.org/10.5209/hics.66290

Allcott, H., & Gentzkow, M. (2017). Social media and *Fake News* in the 2016 election. Journal of Economic Perspectives, 31(2), 211–236. https://doi.org/10.1257/jep.31.2.211

Almutairi, Z., & Elgibreen, H. (2022). A Review of Modern Audio Deepfake Detection Methods: Challenges and Future Directions. *Algorithms*, *15*(5). https://doi.org/10.3390/A15050155

Alvi, I., & Saraswat, N. (2021). Motivation Versus Intention of Sharing *Fake News* Among Social Media Users during the Pandemic-A SEM Model. Journal of Contemporary Eastern Asia, 20(2), 40–62. https://doi.org/10.17477/jcea.2021.20.2.040

Anderson, E. (2017). *Building trust online by partnering with the International Fact Checking Network*.

https://medium.com/google-news-lab/building-trust-online-by-partnering-with-the-international-fact-checking-network-4aeb774ca0ba

Anspach, N. M. (2017). The New Personal Influence: How Our Facebook Friends Influence the News We Read. *Political Communication*, *34*(4), 590–606. https://doi.org/10.1080/10584609.2017.1316329/SUPPL_FILE/UPCP_A_1316329_SM7954.ZIP

Acharya, R. (2019). *SEO Ranking Factors (2019) - Google's 200+ Algorithm Signals*. https://searchenginelaws.com/article/seo-ranking-factors/

AIMC EGM. (2020). Retrieved April 19, 2020, from http://reporting.aimc.es/index.html#/main/internet

Ariely, D. (2008). *Predictably Irrational: The Hidden Forces That Shape Our Decisions (Harper; 2008)*. www.predictablyirrational.com

Asociación para la Investigación de Medios de Comunicación, AIMC (2021). *Marco general de los medios en España*. https://www.aimc.es/a1mc-c0nt3nt/uploads/2021/02/marco2021.pdf

Babu, A., Liu, A., & Zhang, J. (2017). *New Updates to Reduce Clickbait Headlines | Facebook Newsroom*. https://newsroom.fb.com/news/2017/05/news-feed-fyi-new-updates-to-reduce-clickbait-headlines/

Backstrom, L. (2013). *News Feed FYI: A Window Into News Feed*. https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed

Bakshy, E. (2012). *Rethinking Information Diversity in Networks*. https://newsroom.fb.com/news/2012/01/rethinking-information-diversity-in-networks/

Bale, P., Walmsley, D., Wustemann, L., Garcia, M., Ericson, M., & Barber, G. (2018). *Citations & References Trust Indicator - Google Docs*. https://docs.google.com/document/d/10QQuXvle8bT3hqRiGCaato6qhO_UwaM3-bsUdTsO2j0/edit

Bapna, A., & Park, S. (2017). *Updating How We Account For Video Completion Rates | Facebook Newsroom*. https://newsroom.fb.com/news/2017/01/news-feed-fyi-updating-how-we-account-for-video-completion-rates/

Baptista, J. P., & Gradim, A. (2022). A Working Definition of Fake News. *Encyclopedia 2022, Vol. 2, Pages 632-645*, *2*(1), 632–645. https://doi.org/10.3390/ENCYCLOPEDIA2010043

Baptista, J. P., Correia, E., Gradim, A., & Piñeiro-Naval, V. (2021). The influence of political ideology on fake news belief: The Portuguese case. *Publications*, *9*(2). https://doi.org/10.3390/PUBLICATIONS9020023/S1

Bhargava, V. R., & Velasquez, M. (2021). Ethics of the Attention Economy: The Problem of Social Media Addiction. Business Ethics Quarterly, 31(3), 321–359. https://doi.org/10.1017/BEQ.2020.32

Bovet, A., & Makse, H. A. (2019). Influence of *Fake News* in Twitter during the 2016 US presidential election. https://doi.org/10.1038/s41467-018-07761-2

Bradley, T. (2009). *Facebook Makeover: The Good, the Bad, and the Backlash | PCWorld*. https://www.pcworld.com/article/174313/faebook_makeover_good_bad_backlash.html

Bradshaw, S., & Howard, P. N. (2018). Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation. Retrieved from http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf

Bright, L. F., Kleiser, S. B., & Grau, S. L. (2015). Too much Facebook? An exploratory examination of social media fatigue. Computers in Human Behavior, 44, 148–155. https://doi.org/10.1016/J.CHB.2014.11.048

Buchanan, T. (2020). Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. *PLOS ONE*, *15*(10), e0239666. https://doi.org/10.1371/JOURNAL.PONE.0239666

Burkhardt, J. M. (2017). History of *Fake News*. Library Technology Reports, 53(8), 5. https://www.proquest.com/scholarly-journals/history-fake-news/docview/1967322040/se-2?accountid=14501

Burkhardt, J. M. (2017). How *Fake News* Spreads. Library Technology Reports, 53(8), 10–12.

Carden, M. (2018). *Responding to The Guardian: A Fact-Check on Fact-Checking | Meta*. https://about.fb.com/news/2018/12/guardian-fact-check/

Chance, T., Hicks, J., Honderich, R., Srbinovich, M., Lavrov, J., Diversity Working Group, & News Agency Working Group. (2018). *Author/Reporter Expertise Trust Indicator - Google Docs*. https://docs.google.com/document/d/1lK_F_6r__seoE90ze6Kk6geONR3C6OcPmfw7WRJ3GYQ/edit#

Chang, J. (2017). *Identifying credible content online, with help from the Trust Project*. https://www.blog.google/outreach-initiatives/google-news-initiative/sorting-through-information-help-trust-project/

Chesney, R., & Citron, D. K. (2018, October 16). Disinformation on Steroids: The Threat of Deep Fakes. https://www.cfr.org/report/deep-fake-disinformation-steroids

Chesney, B., & Citron, D. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. California Law Review, 107. https://heinonline.org/HOL/Page?handle=hein.journals/calr107&id=1789&div=51&collection=journals

Chowdhry, A. (2015). *Facebook Changes News Feed Algorithm To Prioritize Content From Friends Over Pages*. https://www.forbes.com/sites/amitchowdhry/2015/04/23/facebook-changes-news-feed-algorithm-to-prioritize-content-from-friends-over-pages/#1a6f4b48127a

Coens, J. (2012). *See Posts That Matter to You | Facebook Newsroom*. https://newsroom.fb.com/news/2012/03/see-posts-that-matter-to-you/

Cote, J. (2022). Deepfakes and *Fake News* Pose a Growing Threat to Democracy - News @ Northeastern. https://news.northeastern.edu/2022/04/01/deepfakes-fake-news-threat-democracy/

Couto, C., & Modesto, J. G. (2020). *The influence of Facebook on Political Activism and Radicalism - ProQuest*. Retrieved August 10, 2022, from https://www.proquest.com/docview/2671684772?pq-origsite=primo

Cramer-Flood, E. (2021). *Duopoly still rules the global digital ad market, but Alibaba and Amazon are on the prowl - Insider Intelligence Trends, Forecasts & Statistics*. https://www.insiderintelligence.com/content/duopoly-still-rules-global-digital-ad-market-alibaba-amazon-on-prowl

Crowe, A. (2021). *What Exactly Is E-A-T & Why Does It Matter to Google?* https://www.searchenginejournal.com/google-eat/what-is-it/

D'Onfro, Jillian 2016. 10 years ago Facebook had "the most inglorious launch moment in history" but it changed everything [WWW Document]. Business Insider. URL https://www.businessinsider.com/facebook-news-feed-launch-2016-9 (accessed 5.19.19).

Davey, A. (2020). *Tool to Help Journalists Spot Doctored Images Is Unveiled by Jigsaw - The New York Times*. https://www.nytimes.com/2020/02/04/technology/jigsaw-doctored-images-disinformation.html

David, C. C., San Pascual, R. S., & Torres, E. S. (2019). Reliance on Facebook for news and its influence on political engagement. *PLoS ONE*, *14*(3). https://doi.org/10.1371/JOURNAL.PONE.0212263

Dias, J. A., Doca, H. H., & Silva, F. F. da. (2021). Bots, *Fake News*, fake faces and deepfakes: automation, under the bias of dromology, as a sophisticated form of biopower to influence the democratic election process. Pensar - Revista de Ciências Jurídicas, 26(3), 1–14. https://doi.org/10.5020/2317-2150.2021.11840

Dickinson, B. (2012). *This Is The Social Graph Explained - Business Insider*. http://www.businessinsider.com/explainer-what-exactly-is-the-social-graph-2012-3?IR=T

Dufour, N., & Gully, A. (2019). *Google AI Blog: Contributing Data to Deepfake Detection Research*. https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html

El-Arini, K., & Tang, J. (2014). *Click-baiting | Facebook Newsroom*. https://newsroom.fb.com/news/2014/08/news-feed-fyi-click-baiting/

Elías, C. (2015). *El Selfie de Galileo. Software social, político e intelectual del siglo XXI* (primera). Grup Editorial 62, S.L.U.

Elías, Carlos 2019. *Science on the Ropes. Decline of Scientific Culture in the Era of Fake News*. Springer-Nature. Cham, Switzerland.

Elías, C., & Catalan-Matamoros, D. (2020). Coronavirus in Spain: Fear of 'Official' *Fake News* boosts WhatsApp and alternative sources. Media and Communication, 8(2), 462–466. https://doi.org/10.17645/MAC.V8I2.3217

Elías, C., Teira, D., Fernández-Roldán Díaz, A., González Moreno, D., García Marín, D., Concepción Mateos Martín, M., Pampín Quian, A., Catalán Matamoros, D., Carral Viral, U., Tuñón Navarro, J., & Zamora Bonill, J. P. (2021). *Manual de periodismo y verificación de noticias en la era de las fake news*.

Erskine, Ryan 2018. Facebook Engagement Sharply Drops 50% Over Last 18 Months [WWW Document]. Forbes. URL https://www.forbes.com/sites/ryanerskine/2018/08/13/study-facebook-engagement-sharply-drops-50-over-last-18-months/ (accessed 5.19.19).

Facebook n.d. *Machine Learning*. Facebook Research. URL https://research.fb.com/category/machine-learning/ (accessed 5.19.19).

Facebook. (2007). *Facebook Develops Network Portals, New Inbox and Updates Site Design | Facebook Newsroom*. https://newsroom.fb.com/news/2007/04/facebook-develops-network-portals-new-inbox-and-updates-site-design/

Facebook. (2007). *Facebook Unveils Facebook Ads | Facebook Newsroom*. https://newsroom.fb.com/news/2007/11/facebook-unveils-facebook-ads/

Facebook. (2008). *Facebook Expands Power of Platform Across the Web and Around the World | Facebook Newsroom*. https://newsroom.fb.com/news/2008/07/facebook-expands-power-of-platform-across-the-web-and-around-the-world/

Facebook. (2008). *Facebook Unveils Next Evolution of Site Design | Facebook Newsroom*. https://newsroom.fb.com/news/2008/07/facebook-unveils-next-evolution-of-site-design/

Facebook. (2014). *Introducing Facebook Media | Facebook Newsroom*. https://newsroom.fb.com/news/2014/09/introducing-facebook-media/

Facebook. (2017). *For Video, Intent and Repeat Viewership Matter | Facebook Newsroom*. https://newsroom.fb.com/news/2017/12/news-feed-fyi-for-video-intent-repeat-viewership-matter/

Facebook 2019. Company Info | Facebook Newsroom. URL https://newsroom.fb.com/company-info/ (accessed 5.25.19).

Facebook. (2019). *What Is Facebook Doing to Address the Challenges It Faces? - About Facebook*. https://about.fb.com/news/2019/02/addressing-challenges/

Fitzpatrick, B. (2007). *Brad's Thoughts on the Social Graph*. http://bradfitz.com/social-graph-problem/

Fox, A. (2019). Majority of Americans were not exposed to '*Fake News*' in 2016 U.S. election, Twitter study suggests. Science. https://doi.org/10.1126/SCIENCE.AAW7885

Franck, G. (2019). The economy of attention. Journal of Sociology, 55(1), 8–19. https://doi.org/10.1177/1440783318811778

Garcia, A. (2022). *Nitol, una nueva botnet especializada en ataques DDoS | ALERTAS | CSO España*. https://cso.computerworld.es/alertas/nitol-una-nueva-botnet-especializada-en-ataques-ddos

Geeng, C., Yee, S., & Roesner, F. (2020). *Fake News* on Facebook and Twitter: Investigating How People (Don't) Investigate. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3313831.3376784

Gelfert, A. (2018). *Fake News*: A Definition. Informal Logic, 38(1), 84–117. https://doi.org/10.22329/il.v38i1.5068

George, J., Gerhart, N., & Torres, R. (2021). Uncovering the Truth about *Fake News*: A Research Model Grounded in Multi-Disciplinary Literature. Journal of Management Information Systems, 38(4), 1067–1094. https://doi.org/10.1080/07421222.2021.1990608/SUPPL_FILE/MMIS_A_1990608_SM9674.DOCX

Giansiracusa, N. (2021). How Algorithms Create and Prevent *Fake News* : Exploring the Impacts of Social Media, Deepfakes, GPT-3, and More.

Giansiracusa, N. (2021). Gravitating to Google. How Algorithms Create and Prevent *Fake News*, 119–150. https://doi.org/10.1007/978-1-4842-7155-1_6

Girasa, R. (2020). Artificial intelligence as a disruptive technology: Economic transformation and government regulation. In Artificial Intelligence as a Disruptive Technology: Economic Transformation and Government Regulation. Palgrave Macmillan. https://doi.org/10.1007/978-3-030-35975-1

Gleicher, N. (2019). *Removing Coordinated Inauthentic Behavior from Russia - About Facebook*. https://about.fb.com/news/2019/01/removing-cib-from-russia/

Gleicher, N. (2019). *Taking Down Coordinated Inauthentic Behavior in Indonesia - About Facebook*. https://about.fb.com/news/2019/01/taking-down-coordinated-inauthentic-behavior-in-indonesia/

Google. (2019). *Cómo funciona la Búsqueda de Google | Descripción general*. https://www.google.com/search/howsearchworks/

Google. (2019). *Cómo funciona la Búsqueda de Google | Respuestas útiles*. https://www.google.com/search/howsearchworks/responses/

Google. (2022). *General Guidelines*. https://static.googleusercontent.com/media/guidelines.raterhub.com/en//searchqualityevaluatorguidelines.pdf

Google. (2019). *Google News Lab – Google News Initiative*. https://newsinitiative.withgoogle.com/google-news-lab

Google. (2019). *How Google Search Works - Search Console Help*. https://support.google.com/webmasters/answer/70897?hl=en

Google. (2018). *Introducing the Google News Initiative - YouTube*. https://www.youtube.com/watch?v=d6ihrHNNkkY

Google. (2022). *Extensions - Chrome Developers*. https://developer.chrome.com/docs/extensions/

Google. (2022). *Jigsaw*. https://jigsaw.google.com/

Google. (2020, Febrero 27). *Overview | Search for Developers | Google Developers*. https://developers.google.com/search/reference/overview

Google. (2019). *SyntaxNet – opensource.google.com*. https://opensource.google.com/projects/syntaxnet

Greene, C. M., Nash, R. A., & Murphy, G. (2021). Misremembering Brexit: partisan bias and individual predictors of false memories for fake news stories among Brexit voters. *Memory*, *29*(5), 587–604. https://doi.org/10.1080/09658211.2021.1923754

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). *Fake News* on Twitter during the 2016 U.S. presidential election. Science (American Association for the Advancement of Science), 363(6425), 374–378. https://doi.org/10.1126/science.aau2706

Guess, A. M., Nyhan, B., & Reifler, J. (2020). Exposure to untrustworthy websites in the 2016 US election. Nature Human Behaviour, 4(5), 472–480. https://doi.org/10.1038/S41562-020-0833-X

Gunther, R., Nisbet, E. C., & Beck, P. (2018). Trump may owe his 2016 victory to "*Fake News*," new study suggests. https://theconversation.com/trump-may-owe-his-2016-victory-to-fake-news-new-study-suggests-91538

Hao, K. (2020). *Google has released a tool to spot faked and doctored images | MIT Technology Review*. https://www.technologyreview.com/2020/02/05/349126/google-ai-deepfakes-manipulated-images-jigsaw-assembler/

Hardiman, A., & Brown, C. (2018). *More Local News on Facebook, Globally | Facebook Newsroom*. https://newsroom.fb.com/news/2018/03/news-feed-fyi-more-local-news-on-facebook-globally/

Haucap, J., & Heimeshoff, U. (2014). Google, Facebook, Amazon, eBay: Is the Internet driving competition or market monopolization? International Economics and Economic Policy, 11(1–2), 49–61. https://doi.org/10.1007/s10368-013-0247-6

Hu, M. (2020). Cambridge Analytica's black box. https://doi.org/10.1177/2053951720938091

Hughes, T., Smith, J., & Leavitt, A. (2018). *Helping People Better Assess the Stories They See in News Feed | Facebook Newsroom*. https://newsroom.fb.com/news/2018/04/news-feed-fyi-more-context/

Hsu, E., Magnuson, K., Go, K., Begay, J., Cheung, P., Pihlajamaki, R., & Powell, T. M. (2017). *Diverse Voices Trust Indicator - Google Docs*. https://docs.google.com/document/d/1Fbah0unasIP4GEXVLtxsxVo5u7qSlagIJ4LSfXuMAX8/edit

IBM. (2020, July 2). What is *Machine Learning*? | IBM. https://www.ibm.com/cloud/learn/machine-learning

IBM. (2020, July 2). What is Natural Language Processing? | IBM. https://www.ibm.com/cloud/learn/natural-language-processing

International Telecommunication Union. (2021). *Measuring digital development - Facts and figures 2021*.

Jacobsson, B. (n.d.). *Cybercriminal Organizations : Utilization of Botnets*. Retrieved September 25, 2022, from www.bth.se

Jagpal, N., Dingle, E., Gravel, J.-P., Mavrommatis, P., Provos, N., Rajab, A., Thomas, K., Moheeb, N. P., & Google, K. T. (2015). *Trends and Lessons from Three Years Fighting Malicious Extensions*.

Jigsaw. (2022). *Assembler — a Jigsaw experiment*. https://projectassembler.org/

Jigsaw. (2022). *Data Visualizer | Jigsaw*. https://jigsaw.google.com/the-current/disinformation/dataviz/

Jigsaw. (2022). *Issues | Jigsaw*. https://jigsaw.google.com/issues/#disinformation

Jigsaw. (2022). *Assembler — Learnings*. https://projectassembler.org/learnings/

Jiménez Iglesias, L., Aguilar Paredes, C., Sánchez Gómez, L., & Gutiérrez, M. P. M. (2018). User experience and media. The three click rule in newspapers' webs for smartphones. *Revista Latina de Comunicacion Social*, *73*, 595–613. https://doi.org/10.4185/RLCS-2018-1271

Jung, T., Kim, S., & Kim, K. (2020). DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern. *IEEE Access*, *8*, 83144–83154. https://doi.org/10.1109/ACCESS.2020.2988660

Kaspersky. (2022). *What is a Botnet?* https://www.kaspersky.com/resource-center/threats/botnet-attacks

Kaspersky. (2022). What is adware? https://www.kaspersky.com/resource-center/threats/adware

Keysar, D., & Shmiel, T. (2013). *US9477759B2 - Question answering using entity references in unstructured data - Google Patents* (Patent No. US9477759B2). https://patents.google.com/patent/US9477759

Kozik, R., Kula, S., Choraś, M., & Woźniak, M. (2022). Technical solution to counter potential crime: Text analysis to detect *Fake News* and disinformation. *Journal of Computational Science*, *60*, 101576. https://doi.org/10.1016/J.JOCS.2022.101576

Kreiss, D., & Mcgregor, S. C. (2017). Technology Firms Shape Political Communication: The Work of Microsoft, Facebook, Twitter, and Google With Campaigns During the 2016 U.S. Presidential Cycle. Https://Doi.Org/10.1080/10584609.2017.1364814, 35(2), 155–177. https://doi.org/10.1080/10584609.2017.1364814

Kacholia, V., & Ji, M. (2013). *Helping You Find More News to Talk About | Facebook Newsroom*. https://newsroom.fb.com/news/2013/12/news-feed-fyi-helping-you-find-more-news-to-talk-about/

Kshetri, N., & Voas, J. (2017). The Economics of "*Fake News*." IT Professional, 19(6), 8–12. https://doi.org/10.1109/MITP.2017.4241459

Lada, A., Li, J., & Ding, S. (2017). *New Signals to Show You More Authentic and Timely Stories | Facebook Newsroom*. https://newsroom.fb.com/news/2017/01/news-feed-fyi-new-signals-to-show-you-more-authentic-and-timely-stories/

Lafferty, J. (2013). *Facebook Announces 'Story Bumping,' And Other Small Changes To News Feed Algorithm – Adweek*. https://www.adweek.com/digital/facebook-announces-story-bumping-and-other-small-changes-to-news-feed-algorithm/

Landon-Murray, M., Mujkic, E., & Nussbaum, B. (2019). Disinformation in Contemporary U.S. Foreign Policy: Impacts and Ethics in an Era of *Fake News*, Social Media, and Artificial Intelligence. Public Integrity, 21(5), 512–522. https://doi.org/10.1080/10999922.2019.1613832

Leathern, R., & Chang, B. (2017). *Addressing Cloaking So People See More Authentic Posts | Facebook Newsroom*. https://newsroom.fb.com/news/2017/08/news-feed-fyi-addressing-cloaking-so-people-see-more-authentic-posts/

Levy, C. (2022). Así fue la campaña de propaganda de la mayor empresa de reparto de comida de Brasil para frenar las protestas laborales. https://www.eldiario.es/internacional/campana-propaganda-mayor-empresa-reparto-comida-brasil-frenar-protestas-laborales_1_9104692.html

Li, X., Lu, P., Hu, L., Wang, X., & Lu, L. (2021). A novel self-learning semi-supervised deep learning network to detect *Fake News* on social media. Multimedia Tools and Applications, 81, 19341–19349. https://doi.org/10.1007/s11042-021-11065-x

Lin, J.-R., & Guo, S. (2017). *Reducing Links to Low-Quality Web Page Experiences | Facebook Newsroom*. https://newsroom.fb.com/news/2017/05/reducing-links-to-low-quality-web-page-experiences/

Lyons, K. (2020). *Google is indexing WhatsApp group chat links, making private groups discoverable - The Verge*. https://www.theverge.com/2020/2/21/21147073/whatsapp-google-group-chat-join-indexing-links-search-privacy-facebook

Lyons, T. (2018). *Increasing Our Efforts to Fight False News | Facebook Newsroom*. https://newsroom.fb.com/news/2018/06/increasing-our-efforts-to-fight-false-news/

Lysak, S., Cremedas, M., & Wolf, J. (2012). Facebook and Twitter in the Newsroom. *Electronic News*, *6*(4), 187–207. https://doi.org/10.1177/1931243112466095

Manjoo, Farhad 2013. The Most Influential Feature on the Internet Quietly Launched Seven Years Ago [WWW Document]. Slate Magazine. URL https://slate.com/technology/2013/09/facebook-news-feed-turns-7-why-its-the-most-influential-feature-on-the-internet.html (accessed 5.19.19).

Marc Owen, J. (2019). The Gulf Information War| Propaganda, *Fake News*, and Fake Trends: The Weaponization of Twitter Bots in the Gulf Crisis | Jones | International Journal of Communication. https://ijoc.org/index.php/ijoc/article/view/8994

Marcos Recio, J. C., Edo Bolós, C., & Parra Valcarce, D. (2018). Remaining challenges for digital newspapers regarding informative updates: Case studies in the Spanish media. *Communication and Society*, *31*(2), 51–70. https://doi.org/10.15581/003.31.2.51-70

Masciari, E., Moscato, V., Picariello, A., & Sperlí, G. (2020). Detecting *Fake News* by image analysis. *ACM International Conference Proceeding Series*, *5*(2020). https://doi.org/10.1145/3410566.3410599

Masera, A., Barber, G., Stewart, M., Northrop, P., LeCompte, C., Soldal, H., Deighton, B., Bernabó, R., Smit, M., Maushard, B., Montgomery, J., Griwert, K., Honderich, R., & Smith, A. (2018). *Type of Work Trust Indicator - Google Docs*. https://docs.google.com/document/d/1pLUM-JvGDl5NMUNMi2Hl74MnsoBGw_B3cYIAQ8NyO_M/edit

Massoglia, A. (2020). 'Dark money' networks hide political agendas behind *Fake News* sites • OpenSecrets. https://www.opensecrets.org/news/2020/05/dark-money-networks-fake-news-sites/

Massolo, A., & Traversi, M. (2021). Is it possible to mitigate cognitive biases? A critical analysis of different proposals for reducing myside bias. *Prometeica*, *23*, 60–76. https://doi.org/10.34024/PROMETEICA.2021.23.11419

Mauschard, B., Ackermann, S., Porto, M., & Mungeam, F. (2017). *Local Reporting Trust Indicator - Google Docs*. https://docs.google.com/document/d/1elu7Q94Qf9amlhVDDmVQgGf961auSGSfXlNA1JeW5H4/edit

Mays, L. (2015). The consequences of search bias: how application of the essential facilities doctrine remedies Google's unrestricted monopoly on search in the United States and Europe. The George Washington Law Review, 83(2), 721-.

McCarthy, C. (2010). *Facebook F8: One graph to rule them all - CNET*. https://www.cnet.com/culture/facebook-f8-one-graph-to-rule-them-all/

McLuhan, M. (Ed.). (1996). Comprender los medios de comunicación Las extensiones del ser humano (1st ed.). Paidós. ISBN: 84-493-0240-4

Merriam-Webster. (2019). *Web Crawler | Definition of Web Crawler by Merriam-Webster*. https://www.merriam-webster.com/dictionary/web%20crawler

Miller, D. (2019). *Updates to Video Ranking - About Facebook*. https://about.fb.com/news/2019/05/updates-to-video-ranking/

Milmo, D. (2022, Marzo 4). Russia blocks access to Facebook and Twitter | Russia | The Guardian. https://www.theguardian.com/world/2022/mar/04/russia-completely-blocks-access-to-facebook-and-twitter

Miners, Z. (2014). *Facebook will be mostly video in 5 years, Zuckerberg says | PCWorld*. https://www.pcworld.com/article/2844852/facebook-will-be-mostly-video-in-5-years-zuckerberg-says.html

Mitchell, A. (2014). *Announcing FB Newswire, Powered by Storyful | Facebook Newsroom*. https://newsroom.fb.com/news/2014/04/announcing-fb-newswire-powered-by-storyful/

Mitchell, A. Kiley, J. Gottfried, J. & Guskin, E. 2013. "The Role of News on Facebook". Pew Research Center. Journalism and Media. https://www.journalism.org/2013/10/24/the-role-of-news-on-facebook/

Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2019). "*Fake News*" Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. American Behavioral Scientist, 000276421987822. https://doi.org/10.1177/0002764219878224

Mosseri, A. (2016). *Addressing Hoaxes and Fake News | Facebook Newsroom*. https://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/

Mosseri, A. (2016). *Building a Better News Feed for You | Facebook Newsroom*. https://newsroom.fb.com/news/2016/06/building-a-better-news-feed-for-you/

Mosseri, A. (2017). *Showing More Informative Links in News Feed | Facebook Newsroom*. https://newsroom.fb.com/news/2017/06/news-feed-fyi-showing-more-informative-links-in-news-feed/

Mosseri, A. (2018). *Bringing People Closer Together | Facebook Newsroom*. https://newsroom.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/

Mosseri, A. (2018). *Helping Ensure News on Facebook Is From Trusted Sources | Facebook Newsroom*. https://newsroom.fb.com/news/2018/01/trusted-sources/

Mozilla. (2022). *Introduction to the DOM - Web APIs | MDN*. https://developer.mozilla.org/en-US/docs/Web/API/Document_Object_Model/Introduction

Muraleedharan, S. (2017). *Introducing Snooze to Give You More Control Of Your News Feed | Facebook Newsroom*. https://newsroom.fb.com/news/2017/12/news-feed-fyi-snooze/

Muraleedharan, S. (2018). *Keyword Snooze: A New Way to Help Control Your News Feed | Facebook Newsroom*. https://newsroom.fb.com/news/2018/06/keyword-snooze-a-new-way-to-help-control-your-news-feed/

Myschasuk, E., Garcia-Ruiz, E., Carpenter, C., Davis, P., Roman, E., Drescher, J., & Gingras, R. (2018). *Methods Trust Indicator - Google Docs*. https://docs.google.com/document/d/1Oy1Gbo9MFf4GpeTVSxD-c2tpf6m6odaYNYY1zmL8UGs/edit#

Newton, C. (2019). *A new Facebook News tab is starting to roll out in the United States - The Verge*. https://www.theverge.com/2019/10/25/20930664/facebook-news-tab-launch-united-states-test

Nickerson, R. S. (1998). "Confirmation bias: A ubiquitous phenomenon in many guises". Review of General Psychology. 2(2).175-220.

O'Donnell, L. (2020). *Overlay Malware Exploits Chrome Browser, Targets Banks and Heads to Spain | Threatpost*. https://threatpost.com/overlay-malware-exploits-chrome-browser-targets-banks-and-heads-to-spain/154713/

O'Reilly, L. (2014). *SocialBakers Finds Facebook Videos Overtook YouTube Videos Posted on By Facebook Pages For The First Time - Business Insider*. https://www.businessinsider.com/facebook-video-v-youtube-market-share-data-2014-12?IR=T

Oracle. (2022). ¿Qué es el malware? Definición de malware | Oracle España. https://www.oracle.com/es/database/security/que-es-el-malware.html

Osmundsen, M., Bor, A., Vahlstrup, P. B., Bechmann, A., & Petersen, M. B. (2021). Partisan Polarization Is the Primary Psychological Motivation behind Political *Fake News* Sharing on Twitter. American Political Science Review, 115(3), 999–1015. https://doi.org/10.1017/S0003055421000290

Ostrow, A. (2009). *Facebook Launching New Real-Time Homepage*. https://mashable.com/2009/03/04/facebook-homepage-real-time/?europe=true

Owens, E., & Vickrey, D. (2014). *Showing More Timely Stories from Friends and Pages | Facebook Newsroom*. https://newsroom.fb.com/news/2014/09/news-feed-fyi-showing-more-timely-stories-from-friends-and-pages/

Owens, E., & Weinsberg, U. (2015). *Showing Fewer Hoaxes | Facebook Newsroom*. https://newsroom.fb.com/news/2015/01/news-feed-fyi-showing-fewer-hoaxes/

Papadogiannakis, E., & Kourtellis, N. (2022). Who Funds Misinformation? A Systematic Analysis of the Ad-related Profit Routines of *Fake News* sites.

Parker, P. (2011). *Facebook Adds "Subscribe" Button, Other News Feed Options - Search Engine Land*.

https://searchengineland.com/facebook-adds-subscribe-button-other-news-feed-options-92816

Paterson, T., & Hanley, L. (2020). Political warfare in the digital age: cyber subversion, information operations and 'deep fakes.' Australian Journal of International Affairs, 74(4), 439–454. https://doi.org/10.1080/10357718.2020.1734772

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50. https://doi.org/10.1016/J.COGNITION.2018.06.011

Pensiero, K., Dernbach, C., Barrett, R., McKenzie, R., Jones, B., Daniszewski, J., Martinez de la Serna, C., Jensen, E., & Diversity Working Group Members. (2019). *Best Practices Trust Indicator - Google Docs*. https://docs.google.com/document/d/1jdt4V92XtveciID3TBl79aiwQcYs5uGSDVdN72PGcpw/edit

Pérez-Escoda, A., Pedrero-Esteban, L. M., Rubio-Romero, J., & Jiménez-Narros, C. (2021). *Fake News* reaching young people on social networks: Distrust challenging media literacy. Publications, 9(2). https://doi.org/10.3390/PUBLICATIONS9020024

Peterson, J. (2022). *'Netflix Party' and Four Other Chrome Extensions That Are Actually Malware*. https://lifehacker.com/netflix-party-and-four-other-chrome-extensions-that-a-1849479234

Peterson, T. (2012). *Another Agency Claims Facebook Algorithm Changes*. https://www.adweek.com/digital/another-agency-claims-facebook-algorithm-changes-144405/

Petrov, S. (2016). *Google AI Blog: Announcing SyntaxNet: The World's Most Accurate Parser Goes Open Source*. https://ai.googleblog.com/2016/05/announcing-syntaxnet-worlds-most.html

Peysakhovich, A. (2016). *Further Reducing Clickbait in Feed | Facebook Newsroom*. https://newsroom.fb.com/news/2016/08/news-feed-fyi-further-reducing-clickbait-in-feed/

Project Veritas. (2019). *SS1DocDump.pdf*. https://www.projectveritas.com/wp-content/uploads/2019/06/SS1DocDump.pdf

Popiołek, M., Hapek, M., & Barańska, M. (2021). Infodemia – an analysis of *Fake News* in polish news portals and traditional media during the coronavirus pandemic. Communication and Society, 34(4), 81–98. https://doi.org/10.15581/003.34.4.81-98

Protalinski, E. (2012). *Facebook starts displaying ads in the News Feed | ZDNet*. https://www.zdnet.com/article/facebook-starts-displaying-ads-in-the-news-feed/

Pu, W., Hu, J., Wang, X., Li, Y., Hu, S., Zhu, B., Song, R., Song, Q., Wu, X., & Lyu, S. (2022). Learning a deep dual-level network for robust DeepFake detection. *Pattern Recognition*, *130*, 108832. https://doi.org/10.1016/J.PATCOG.2022.108832

Przybylski, A. K., Murayama, K., Dehaan, C. R., & Gladwell, V. (2013). Motivational, emotional, and behavioral correlates of fear of missing out. Computers in Human Behavior, 29(4), 1841–1848. https://doi.org/10.1016/J.CHB.2013.02.014

Ralston, S., Kliestik, T., Rowland, Z., & Vrbka, J. (2018). Are pervasive systems of *Fake News* provision sowing confusion? The role of digital media platforms in the production and consumption of factually dubious content. Geopolitics, History, and International Relations, 10(2), 30–36. https://doi.org/10.22381/GHIR10220183

Rait, Z. (2011). *Introducing the Subscribe Button | Facebook Newsroom*. https://newsroom.fb.com/news/2011/09/introducing-the-subscribe-button/

Ravettino Destefanis, A. (2019). El hipertexto: ¿revolución tecnológica o cultural? Cambios y continuidades en la producción y consumo de contenidos textuales. *Revista Internacional de Tecnología, Conocimiento y Sociedad*, *8*(1), 1–8. https://doi.org/10.18848/2474-588X/CGP/V08I01/1-8

Ribeiro, F. N., Saha, K., Babaei, M., Henrique, L., Messias, J., Benevenuto, F., Gummadi, K. P., Redmiles, E. M., & Goga, O. (2019). On Microtargeting Socially Divisive Ads:[0.1em] A Case Study of Russia-Linked Ad Campaigns on Facebook. Proceedings of the Conference on Fairness, Accountability, and Transparency. https://doi.org/10.1145/3287560

Rosen, G. (2019). *Protecting Facebook Live From Abuse and Investing in Manipulated Media Research - About Facebook*. https://about.fb.com/news/2019/05/protecting-live-from-abuse/

Sahoo, S. R., & Gupta, B. B. (2021). Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing*, *100*. https://doi.org/10.1016/J.ASOC.2020.106983

Schema. (2019). *about page - schema.org*. https://schema.org/docs/about.html

Schultz, E. (2003). Pandora's Box: spyware, adware, autoexecution, and NGSCB. Computers & Security, 22(5), 366–367. https://doi.org/10.1016/S0167-4048(03)00501-7

Sethuraman, R. (2019). *Why Am I Seeing This? We Have an Answer for You - About Facebook*. https://about.fb.com/news/2019/03/why-am-i-seeing-this/

Sethuraman, R., Vallmitjana, J., & Levin, J. (n.d.). *Using Surveys to Make News Feed More Personal - About Facebook*. 2019. Retrieved May 10, 2020, from https://about.fb.com/news/2019/05/more-personalized-experiences/

Shahid, W., Li, Y., Staples, D., Amin, G., Hakak, S., & Ghorbani, A. (2022). Are You a Cyborg, Bot or Human?-A Survey on Detecting *Fake News* Spreaders. *IEEE Access*, *10*, 27069–27083. https://doi.org/10.1109/ACCESS.2022.3157724

Shrestha, A., & Spezzano, F. (2021). Textual Characteristics of News Title and Body to Detect *Fake News*: A Reproducibility Study. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 12657 LNCS, 120–133. https://doi.org/10.1007/978-3-030-72240-1_9/TABLES/6

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). *Fake News* Detection on Social Media: A Data Mining Perspective. http://www.journalism.org/2016/05/26/news-use-across-

Shu, K., Wang, S., Lee, D., & Liuu, H. (2020). Lecture Notes in Social Networks Disinformation, Misinformation, and *Fake News* in Social Media Emerging Research

Challenges and Opportunities. Retrieved August 5, 2021, from http://www.springer.com/series/8768

Simo, F. (2014). *The Latest on Facebook Video | Facebook Newsroom*. https://newsroom.fb.com/news/2014/09/the-latest-on-facebook-video/

Silverman, C. (2016). This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook. Retrieved February 26, 2022, from https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook

Silverman, H., & Huang, L. (2017). *Fighting Engagement Bait on Facebook | Facebook Newsroom*. https://newsroom.fb.com/news/2017/12/news-feed-fyi-fighting-engagement-bait-on-facebook/

SimilarWeb. (2020). Google.com Analytics - Market Share Stats & Traffic Ranking. https://www.similarweb.com/website/google.com

Slodkowski, A. (2018). *Facebook bans Myanmar army chief, others in unprecedented move - Reuters*. https://www.reuters.com/article/us-myanmar-facebook/facebook-bans-myanmar-army-chief-others-in-unprecedented-move-idUSKCN1LC0R7

Smarty, A. (2009). *How Google May (Theoretically) Discover Web Pages - Search Engine Journal*. https://www.searchenginejournal.com/google-discover-web-pages/10320/#close

Smyrnaios, N. (2019). Google as an Information Monopoly. Contemporary French and Francophone Studies, 23(4), 442–446. https://doi.org/10.1080/17409292.2019.1718980

Statista. (May 25, 2021). Forecast of the smartphone penetration in Europe from 2010 to 2025 [Graph]. In Statista. Retrieved August 31, 2022, from https://www.statista.com/forecasts/1147144/smartphone-penetration-forecast-in-europe

Song, A. (2014). *Showing Stories About Topics You Like | Facebook Newsroom*. https://newsroom.fb.com/news/2014/02/news-feed-fyi-showing-stories-about-topics-you-like/

Southwell, B. G., Thorson, E. A., & Sheble, L. (2017). The Persistence and Peril of Misinformation. American Scientist, 105(6), 372. https://doi.org/10.1511/2017.105.6.372

Stanford NLP Group. (2019). *The Stanford Natural Language Processing Group*. https://nlp.stanford.edu/software/lex-parser.shtml

Struhar, C. (2014). *Finding Popular Conversations on Facebook | Facebook Newsroom*. https://newsroom.fb.com/news/2014/01/finding-popular-conversations-on-facebook/

Su, S. (2017). *New Test With Related Articles | Facebook Newsroom*. https://newsroom.fb.com/news/2017/04/news-feed-fyi-new-test-with-related-articles/

Sullivan, D., & Illyes, G. (2019). *Official Google Webmaster Central Blog: Evolving "nofollow" – new ways to identify the nature of links*. https://webmasters.googleblog.com/2019/09/evolving-nofollow-new-ways-to-identify.html

Talwar, S., Dhir, A., Kaur, P., Zafar, N., & Alrasheedy, M. (2019). Why do people share *Fake News*? Associations between the dark side of social media use and *Fake News* sharing behavior. Journal of Retailing and Consumer Services, 51, 72–82. https://doi.org/10.1016/J.JRETCONSER.2019.05.026

Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "*Fake News*": A typology of scholarly definitions. Digital Journalism, 6(2), 137–153. https://doi.org/10.1080/21670811.2017.1360143

Tandoc, E. C., Thomas, R. J., & Bishop, L. (2021). What is (Fake) news? analyzing news values (and more) in fake stories. Media and Communication, 9(1), 110–119. https://doi.org/10.17645/MAC.V9I1.3331

Tas, S., & Chiraphadhanakul, T. V. (2015). *Using Surveys to Better Understand Viral Stories | Facebook Newsroom*.

https://newsroom.fb.com/news/2015/12/news-feed-fyi-using-surveys-to-better-understand-viral-stories/

Tchakounté, F., Faissal, A., Atemkeng, M., & Ntyam, A. (2020). A reliable weighting scheme for the aggregation of crowd intelligence to detect *Fake News*. *Information (Switzerland)*, *11*(6), 319. https://doi.org/10.3390/INFO11060319

Terhaar, J., Haymarket, M. P., & Koon, B. (2017). *Actionable Feedback Trust Indicator - Google Docs*. https://docs.google.com/document/d/1lRMVKh_93QuzyW8ce5BxTusLszzzVMyqJ39BL9QTAlM/edit#heading=h.5j7dqc3b2xoj

The Trust Project. (2019). *Collaborator Materials.* Retrieved August 16, 2019, from https://thetrustproject.org/collaborator-materials/

The Trust Project. (2019). *Frequently Asked Questions*. Retrieved September 1, 2019, from https://thetrustproject.org/faq/

The Trust Project. (2019). *The Trust Project*. Retrieved September 1, 2019, from https://thetrustproject.org/

Tonkelowitz, M. (2011). *Interesting News, Any Time You Visit | Facebook Newsroom*. https://newsroom.fb.com/news/2011/09/interesting-news-any-time-you-visit/

Törnberg, P. (2018). Echo chambers and viral misinformation: Modeling *Fake News* as complex contagion. PLOS ONE, 13(9), e0203958. https://doi.org/10.1371/JOURNAL.PONE.0203958

Tosswill, C. (2015). *How the Reactions Test Will Impact Ranking | Facebook Newsroom*. https://newsroom.fb.com/news/2015/10/news-feed-fyi-how-the-reactions-test-will-impact-ranking/

Trninić, D., Vukelić, A. K., & Bokan, J. (2022). Perception of "*Fake News*" and potentially manipulative content in digital media—a generational approach. Societies, 12(1). https://doi.org/10.3390/SOC12010003

Urban, T., Tatang, D., Holz, T., & Pohlmann, N. (2018). Towards Understanding Privacy Implications of Adware and Potentially Unwanted Programs. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11098 LNCS*, 449–469. https://doi.org/10.1007/978-3-319-99073-6_22

van Grove, J. (2009). *3 Key Reasons Facebook Bought FriendFeed*. https://mashable.com/2009/08/10/reasons-facebook-friendfeed/?europe=true#SeO6O7aol8ql

Veritone. (2022). Everything You Need To Know About Deepfake Voice | Veritone. https://www.veritone.com/blog/everything-you-need-to-know-about-deepfake-voice/

Wang, M., & Zhuo, Y. (2015). *Taking into Account More Actions on Videos | Facebook Newsroom*. https://newsroom.fb.com/news/2015/06/news-feed-fyi-taking-into-account-more-actions-on-videos/

Warman, M. (2013). *Facebook hashtags #introduced - Telegraph*. https://www.telegraph.co.uk/technology/facebook/10117483/Facebook-hashtags-introduced.html

Wasserman, T. (2013). *Facebook Tweaks Algorithm for Ads in the News Feed*. https://mashable.com/2013/09/27/facebook-algorithm-ads/?europe=true#UrNkNGBRagqc

Watson, C. A. (2018). Information Literacy in a Fake/False News World: An Overview of the Characteristics of Fake News and its Historical Development. *International Journal of Legal Information*, *46*(2), 93–96. https://doi.org/10.1017/JLI.2018.25

Webopedia. (2020). *What is Structured Data?* https://www.webopedia.com/TERM/S/structured_data.html

Welch, B., & Zhang, X. (2014). *Showing Better Videos | Facebook Newsroom*. https://newsroom.fb.com/news/2014/06/news-feed-fyi-showing-better-videos/

Whittaker, J. (2020). Tech Giants, Artificial Intelligence, and the Future of Journalism. Retrieved February 26, 2022, from https://www.routledge

Widman, J. (n.d.). *EdgeRank*. Retrieved July 15, 2018, from http://edgerank.net/

Wisker, Z. L., & McKie, R. N. (2021). The effect of *Fake News* on anger and negative word-of-mouth: moderating roles of religiosity and conservatism. Journal of Marketing Analytics, 9(2), 144–153. https://doi.org/10.1057/S41270-020-00101-8

Woodford, A. (2019). *The Hunt for False News: EU Edition - About Facebook*. https://about.fb.com/news/2019/04/the-hunt-for-false-news-eu-edition/

Xu, K., Wang, F., Wang, H., & Yang, B. (2020). Detecting *Fake News* over online social media via domain reputations and content understanding. *Tsinghua Science and Technology*, *25*(1), 20–27. https://doi.org/10.26599/TST.2018.9010139

Yu, A., & Tas, S. (2015). *Taking Into Account Time Spent on Stories | Facebook Newsroom*. https://newsroom.fb.com/news/2015/06/news-feed-fyi-taking-into-account-time-spent-on-stories/

Zaini, A., & Zainal, A. (2018). Exploiting DOM Mutation for the Detection of Ad-injecting Browser Extension. *Advances in Intelligent Systems and Computing*, *843*, 657–669. https://doi.org/10.1007/978-3-319-99007-1_61

Zhang, C., & Chen, S. (2016). *Using Qualitative Feedback to Show Relevant Stories | Facebook Newsroom*. https://newsroom.fb.com/news/2016/02/news-feed-fyi-using-qualitative-feedback-to-show-relevant-stories/

Zigmond, D. (2018). *Machine Learning, Fact-Checkers and the Fight Against False News - About Facebook*. https://about.fb.com/news/2018/04/inside-feed-misinformation-zigmond/