

Las redes sociales a juicio. Patrones adictivos, datos y responsabilidad en la economía de la atención*

Social networks on trial. Addictive patterns, data and responsibility in the attention economy.

CARMEN HERRERO SUÁREZ

*Profesora Titular de Derecho Mercantil
Universidad de Valladolid*

Recibido: 03.02.2025 / Aceptado: 06.02.2025

DOI: 10.20318/cdt.2025.9338

Resumen: El surgimiento de las grandes plataformas tecnológicas de redes sociales ha tenido un profundo impacto en la comunicación e interacciones sociales. Junto a sus innegables beneficios, también se han identificado una serie de riesgos o eventuales efectos negativos asociados a las mismas. Si bien, inicialmente, la faceta negativa de las redes sociales se centró en el uso indebido de datos personales y la lesión del derecho de privacidad de sus usuarios, en los últimos años otras cuestiones, especialmente los riesgos de adicción y daños a la salud mental están entrando en el debate sobre sus peligros. El objeto de este trabajo es el análisis crítico de dicho riesgo y las posibilidades de actuación desde el ordenamiento jurídico para solventarlo o mitigarlo. Para ello, se parte del examen y valoración de las distintas acciones judiciales emprendidas contra las principales redes sociales, especialmente en EE.UU., conocidas como las *social media addiction lawsuits*, examinando su viabilidad y la posibilidad de traslación a nuestro ordenamiento jurídico.

Palabras clave: Redes sociales, adicción, daños, economía de la atención, algoritmos.

Abstract: The emergence of the large social networking technology platforms has had a profound impact on communication and social interaction. Alongside their undeniable benefits, a series of risks or possible negative effects associated with them have also been identified. Although initially the negative side of social networks focused on the misuse of personal data and the violation of their users' right to privacy, in recent years other issues, especially the risks of addiction and damage to mental health, are entering the debate on their dangers. The aim of this work is to critically analyse this risk and the possibilities for action from the legal system to solve or mitigate it. To this end, it starts from the examination and assessment of the various legal actions taken against the main social networks, especially in the USA, known as the *social media addiction lawsuits*, examining their viability and the possibility of transferring them to our legal system.

Keywords: social networks, addiction, damages, attention economy, algorithm.

Sumario: I. Los claroscuros de las redes sociales. II. La economía de la atención y las redes sociales. 1. La generación y monetización de la atención. 2. Las principales demandadas: Meta (Facebook, Instagram), Alphabet (YouTube); Bytedance (TikTok). III. Acciones judiciales contra las

*Este trabajo se ha llevado a cabo en el marco del Proyecto de Investigación: "La consolidación de la revolución digital y sus desafíos: adaptación de las estrategias competitivas de los operadores económicos y de las normas ordenadoras de los mercados" del Ministerio de Ciencia e Innovación. PID2022- 139741OB-C21.

redes sociales. 1. De la privacidad a la adicción. 2. Las demandas sobre adicción a las redes sociales en EE.UU. (*social media addiction lawsuits*). A) Daños derivados del contenido. B) Daños derivados del carácter adictivo de la plataforma. C) Daños derivados del engaño: ocultamiento del carácter adictivo de la plataforma (*failure to warn claims*). IV. ¿Y el futuro? Armas existentes y soluciones regulatorias. 1. *Nihil novum sub sole*. 2. Reflexión final.

I. Los clarososcuros de las redes sociales

1. Uno de los ámbitos o sectores en los que la revolución tecnológica digital de las últimas décadas ha tenido un efecto más evidente, ha sido el de la comunicación y las interacciones sociales, con el surgimiento de plataformas de muy distinta naturaleza, las redes sociales o *social networks*, que, desde distintos formatos, exploran las nuevas posibilidades relacionales. Las redes sociales han tenido un profundo impacto en las relaciones profesionales y personales, permitiendo la conexión instantánea, el intercambio de ideas y la colaboración en proyectos, entre sujetos de diferentes partes del mundo.

2. Una red social digital puede ser definida como “una plataforma de interacción on line que permite al usuario, de forma completamente descentralizada, generar un perfil público, compartir información, colaborar en la generación de contenidos y participar de forma espontánea en movimientos sociales y corrientes de opinión¹”.

3. Las relaciones se producen entre usuarios, horizontalmente, y se establecen a través del intercambio de mensajes, fotos, vídeos, y todo tipo de archivos, en definitiva, de información y datos. Todas estas características son comunes a las redes sociales existentes, sin perjuicio de que, atendiendo al perfil del usuario o a la función de la red, podamos distinguir redes generalistas, las más extendidas, como Facebook, Instagram, Twitter, o bien específicas, como las profesionales o académicas, por ejemplo, LinkedIn o de contactos y citas, como Tinder².

4. Pese a que, en los primeros momentos de aparición de estas nuevas formas de comunicación social, sí se produjo una inicial renovación competitiva y sustitución de distintas plataformas sociales en los mercados (como atestigua, por ejemplo, el auge y caída de Myspace), actualmente, y sin perjuicio de la existencia de numerosas redes sociales de menor cuota, cuatro o cinco plataformas dominan de forma más o menos estable los mercados relacionales³. En este sentido, las redes sociales más usadas en el mundo son: Facebook (3.049 millones); Whatsapp (2.000 millones); You Tube (2.491 millones); Instagram (2.000 millones) y las aplicaciones chinas Wechat (1309 millones) y TikTok (1542 millones usuarios)⁴.

5. El uso de las nuevas tecnologías en comunicación ha aportado grandes beneficios, destacando la mejora de la comunicación interpersonal y el acceso y aprendizaje de nuevos conocimientos. Las redes proporcionan acceso a nuevas formas de contenido, productos y servicios, permiten la interacción

¹ Vid. MARTÍNEZ, E./PACHECHO, M./ATAURI, D.: “Redes sociales digitales: propuestas para una protección jurídica de sus usuarios en España”, *Anuario electrónico de estudios en Comunicación social*, vol.4, núm.2, 2011, p. 147.

² Sobre los distintos tipos de redes sociales, vid. MARTÍNEZ, E./PACHECHO, M./ATAURI, D., *cit.*, pp. 148 y ss.

³ La consolidación de las grandes plataformas no solo es fruto del crecimiento interno empresarial sino también, de una miríada de operaciones de concentración y reestructuración. Muchas de las transacciones realizadas en los mercados digitales obedecen a un patrón similar: concentraciones por las que una gran empresa asentada en el mercado adquiere una empresa recién llegada o *start up*, que ofrece un producto o servicio novedoso, pagándose por ella un precio muy superior al que podría deducirse de su volumen de ingresos. Se trata de operaciones en las que el valor real de la empresa adquirida reside en el carácter innovador de los productos o servicios ofrecidos, la información o datos que controle, así como su presencia en el mercado en términos de número de usuarios de sus servicios. Este patrón ha estado especialmente presente en los mercados de la comunicación digital y en la formación de las grandes plataformas de la atención, por ejemplo, en las operaciones de adquisición de Instagram y Whatsapp por parte de Facebook, LinkedIn por Microsoft o YouTube por Google.

⁴ Datos extraídos de la página Web de Telefónica, julio de 2024: <https://www.telefonica.com/es/sala-comunicacion/blog/redes-sociales-mas-usadas/>

personal en tiempo real y la participación colectiva y ofrecen intermediación entre usuarios y entre grupos de usuarios y empresas.

6. Ahora bien, la irrupción y consolidación de estas nuevas formas de comunicación, especialmente tras el surgimiento de las webs 2.0, que permiten a cualquier persona compartir contenidos e información en red, independientemente de sus conocimientos en programación, no está exenta de polémica y, junto a estos beneficios, también se han puesto de manifiesto una serie de riesgos o eventuales efectos negativos. Inicial y mayoritariamente, la faceta negativa de las redes sociales se ha venido centrando en el uso indebido de datos personales y la posible lesión del derecho a la privacidad de los consumidores o clientes.

7. Pero su cara oscura no se limita a la lesión de la privacidad. En los últimos años y a medida que se ha ido incrementando el uso de las redes sociales, otras cuestiones, especialmente el riesgo de adicción y daños a la salud mental están entrando en el debate sobre sus peligros⁵.

8. Este riesgo se ha puesto de manifiesto a medida que ha ido evolucionando el sistema y los modelos de negocios se han orientado a la búsqueda de un incremento progresivo del compromiso y conexión de los usuarios a fin de maximizar el beneficio de las plataformas. Una de las formas más efectivas para que las plataformas maximicen su valor es mediante el uso de sistemas de recomendación, es decir, sistemas de software sofisticados que utilizan inteligencia artificial para aprender sobre el usuario y predecir qué tipo de contenido los mantendrá conectados. El uso de algoritmos de inteligencia artificial por parte de la plataforma, diseñados fundamentalmente para maximizar las métricas de participación y los daños que pueden derivarse de estos sistemas constituyen el eje central del debate sobre los riesgos actuales de las redes sociales.

9. Un punto de inflexión se produjo con el famoso asunto de los *Facebook Files*, en septiembre de 2021, cuando una empleada de Meta, Frances Haugen, filtró cientos de documentos internos de la compañía que revelaron que sus ejecutivos eran conscientes de los daños que causaban Instagram y Facebook entre los jóvenes, así como de su responsabilidad en la difusión de noticias falsas y de contenidos que alentaban la violencia y ninguna solución o medida real había sido adoptada para prevenirlos⁶.

10. La alarma social generada por esta revelación, unida a diversos casos de daños psicológicos o físicos en niños y adolescentes como consecuencia de su exposición a contenidos peligrosos en las redes, ha determinado la apertura de un debate sobre los riesgos y peligros, especialmente para la salud de los menores de edad, de las redes sociales, así como que estas no hayan escapado al escrutinio y ata-

⁵ Basta repasar los titulares de los principales medios de comunicación de los últimos meses para percibir la existencia de un conflicto social relacionado con los posibles efectos perjudiciales de las redes sociales, especialmente en relación con menores de edad. Así, *inter alia*: “Australia presenta un proyecto de ley “histórico” para prohibir las redes sociales a los menores de 16 años” (El País, 21 de noviembre de 2024); “Las nuevas tabaquerías: la oleada de demandas que puede acabar con las redes sociales” (El País, 22 de abril de 2023); “Tienen sangre en sus manos, los senadores de EE.UU. arremeten contra las redes sociales por no proteger a los menores” (El Economista, 1 de febrero de 2024), “Los retos virales más peligrosos de Tik Tok” (La Vanguardia, 29 de agosto de 2024).

⁶ Bajo el título de *Facebook Files* o *Archivos de Facebook*, el Wall Street Journal, publicó una serie de reportajes, iniciados el 13 de septiembre de 2021, basados en los documentos filtrados, en los que se pone claramente de manifiesto las discrepancias entre la política y afirmaciones públicas de la empresa sobre la seguridad de sus redes sociales y el conocimiento real por parte de sus empleados y directivos sobre los daños y efectos perjudiciales de estas, especialmente en relación con adolescentes y menores de edad. Basta recoger algunas citas literales de los documentos para reflejar esa dicotomía: “We make body image issues worse for one in three teen girls” (Empeoramos los problemas de imagen corporal de una de cada tres adolescentes); “We are not actually doing what we say we do publicly” (En realidad, no hacemos lo que decimos públicamente.), “Misinformation, toxicity and violent content are inordinately prevalent among reshares” (La información errónea, la toxicidad y el contenido violento prevalecen desmesuradamente entre las reacciones compartidas); “An unfortunate side effect is that harmful and misinformative content can go viral, often before we can catch it and mitigate its effects” (un efecto secundario desafortunado es que el contenido dañino y desinformativo puede volverse viral, a menudo antes de que podamos detectarlo y mitigar sus efectos). Vid. <https://www.wsj.com/articles/the-facebook-files-11631713039>.

que que, desde diversos sectores del ordenamiento jurídico, como el Derecho *antitrust*, las normas sobre competencia desleal y daños o el Derecho regulatorio han experimentado las principales plataformas tecnológicas en otros mercados.

11. Pero la lucha judicial no es la única vía de actuación para intentar solventar los problemas de salud y seguridad en la red. Esta va a ser simultaneada con una intervención regulatoria normativa en diversos niveles y con distinto alcance que, con mayor o menor permisividad, va a intentar atajar los riesgos percibidos en los entornos de la comunicación digital mediante la imposición de restricciones y normas de conducta.

12. Si bien las acciones judiciales se están extendiendo frente a todo tipo de redes sociales, tanto generales como especiales, y en numerosas jurisdicciones, destacan por su magnitud y alcance, las acciones emprendidas en EE.UU., fundamentalmente contra las grandes social networks: Facebook, Instagram, YouTube y TikTok, con cientos de demandas interpuestas por particulares, centros educativos, asociaciones de tutela de los consumidores, organismos administrativos y Estados.

13. Antes de exponer e intentar ordenar este *mare magnum* procesal para poder entrar a valorar las bases de las distintas demandas y analizar su viabilidad, así como su posibilidad de transposición a nuestro ordenamiento jurídico, es conveniente conocer un poco más la naturaleza de estas redes sociales y los elementos que estructuran su modelo de negocios, lo que aconseja introducir algunos conceptos interrelacionados, como son los conceptos de economía de la atención y de economía de la vigilancia o *surveillance*.

II. La economía de la atención y las redes sociales

1. La generación y monetización de la atención

14. La “surveillance economy” o economía de la vigilancia se refiere a un modelo económico en el que las empresas recopilan, analizan y utilizan grandes cantidades de datos sobre los usuarios para generar ingresos, lo que incluye la recolección de información personal y la creación de perfiles para dirigir publicidad o contenido específico.

15. Por otro lado, la economía de la atención se centra en cómo se compete por la atención de los usuarios en un entorno saturado de información. El objetivo es la captación y mantenimiento de la atención de las personas, dado que esta constituye un recurso muy valioso en el mundo digital. Desde una perspectiva amplia, comprende aquellas actividades empresariales, especialmente de empresas que basan su negocio en la publicidad, por las que tratan de maximizar el tiempo y la atención que los usuarios prestan a sus productos o servicios (*extract attention for profit*).

16. En resumen, mientras que la economía de la vigilancia se enfoca en la recopilación y uso de datos, la economía de la atención se centra en cómo atraer y retener la atención de los usuarios. Si bien su foco se sitúa en distintos aspectos del compartamiento, tienen su punto de conexión en el hecho de que los datos posibilitan que la atención humana sea computable y monetizable como un activo, en el sentido de que permiten la metrificación del compromiso y la venta de promesas de precisión predictiva a los anunciantes. La economía de plataformas de atención está organizada en torno a una lógica bidireccional que busca tanto aumentar el compromiso como el suministro de atención para producir conocimientos informativos monetizables, y aumentar la captura de datos con el fin de optimizar un compromiso y atención más delgados y fácilmente monetizables. Como señala BIETTI los modelos empresariales de la economía de la atención “se nutren de globos oculares y de datos para obtener un beneficio”⁷.

⁷ Vid. E. BIETTI: “The Data-Attention Imperative”, 2024, Northeastern University School of Law Research Paper No. 473, disponible en: SSRN: <https://ssrn.com/abstract=4729500>, p. 10.

17. Estas activades no son exclusivas de los mercados digitales⁸, si bien las especiales características de estos facilitan el desarrollo de modelos empresariales centrados en la atención y en la recopilación y tratamiento de datos y han propiciado el surgimiento de las grandes plataformas tecnológicas de la atención o la información que pueden describirse como intermediarios que utilizan métodos algorítmicos basados en datos y protocolos de interconexión modulares estandarizados para facilitar las interacciones y transacciones en red digital entre sus usuarios⁹. Entre estas plataformas digitales cuyo modelo de negocio se centra en la captación y monetización de la atención humana, ocupan un papel central las redes sociales¹⁰. Veamos más detenidamente cómo funciona el modelo.

18. Principalmente dos son los elementos que estructuran el sistema: 1) Generación de atención o compromiso (*engagement*) y 2) la monetización de ese compromiso.

19. Por lo que a la **generación de atención** se refiere, la sostenibilidad comercial de las plataformas de atención se basa en mantener altos niveles de compromiso de los usuarios. Estas plataformas utilizan diferentes métodos para captar la atención de los usuarios y generar una dependencia, aprovechando mecanismos infraestructurales, legales y sociales para atraer a nuevos usuarios.

20. Una forma de captación inicial de la atención es mediante el recurso al modelo *freemium* o de subsidio cruzado. El modelo de negocio de estas empresas se encuadra dentro de lo que se conoce como mercados de doble o múltiple cara (*two-sided* o *multisided markets*). Estos mercados se caracterizan por la existencia de una empresa-plataforma que pone en contacto a dos colectivos distintos, vendiendo u ofreciendo dos o más productos o servicios a los dos grupos de consumidores, con demandas relacionadas indirectamente. Una de las características principales de estos mercados es la interdependencia entre ambos lados del mercado, lo que puede llevar a la plataforma a optar por estrategias de fijación de precios que permitan subsidiar el lado del mercado más sensible al precio por parte del lado más sensible al tamaño del otro lado del mercado. La plataforma para maximizar beneficios puede optar así, por ofrecer el servicio de forma gratuita para obtener mayores ingresos en la otra parte del mercado en base a la mayor información recopilada como consecuencia de la gratuidad del servicio. Si el acceso a la red social se ofrece de forma gratuita, es más fácil captar la atención inicial y posteriormente monetizar esa atención transfiriendo la información resultante a anunciantes y otras empresas.

21. La integración vertical y radial de las plataformas, ya sea estructural o contractualmente, posibilita la formación de ecosistemas digitales¹¹ en los que se potencia el compromiso de los usuarios al facilitarles que encuentren todos los servicios que desean en la propia plataforma, eliminando los incentivos para recurrir a proveedores alternativos. Esta misma finalidad puede buscarse mediante otros

⁸ Así, por ejemplo, a este esquema de captación y monetización de la atención responden los modelos empresariales de las cadenas de televisión financiadas mediante la venta de espacios publicitarios. T. WU: analiza la evolución de las empresas de la atención, que el denomina “mercaderes de la atención”, desde la aparición de los primeros periódicos en Nueva York en el s.XIX financiados exclusivamente con publicidad. T. WU: *The attention Merchants*, Atlantic Books, Londres, 2017.

⁹ Vid. E. BIETTI: “The Data-Attention Imperative”, *cit.*, p.10; N. RICHARDS/W. HARTZOG: “Against engament”, *Boston University Law Review*, vol. 104, 2024, pp. 1151 y ss.

¹⁰ E. BIETTI: “The Data-Attention Imperative”, *cit.*, p.10, incluye también dentro de la categoría de plataformas de la atención, a las plataformas de video a la carta, como Netflix o Hulu, y las plataformas de juegos, como, por ejemplo, Steam, Facebook gaming o Twitch.

¹¹ El crecimiento y el correspondiente incremento del grado de concentración de los mercados digitales se caracteriza por su carácter radical. Desde un negocio principal o central la actividad de las empresas se ha ido extendiendo a otros productos o servicios más o menos interrelacionados, dando lugar a auténticos ecosistemas digitales y a la reaparición en los mercados de los grandes conglomerados. En este sentido, por ejemplo, Google no es sólo un mero proveedor de un buscador digital. Ha crecido hasta ofrecer servicios como programación y gestión del tiempo, almacenamiento en la nube, mensajería instantánea, video chat, mapas y navegación, teléfonos inteligentes, etc. y también participa activamente en inteligencia artificial y salud y Meta, de su negocio principal, redes sociales y publicidad, se ha ido extendiendo a otros mercados como la inteligencia artificial, la computación en la nube, la realidad aumentada y el Internet de las cosas. Vid. in extenso, sobre la formación de los grandes ecosistemas digitales, HERRERO SUÁREZ: “Los nuevos conglomerados digitales”, *Cuadernos de Derecho Transnacional (CDT)*, vol.16, núm.2, *Studia Amicorum Alfonso Luis Calvo Caravaca*, 2024. pp. 949 a 965.

mecanismos, como la celebración de acuerdos contractuales con otros actores para incorporar los propios servicios como predeterminados (*default*), lo que los hace complicados y difíciles de sortear.

22. El objetivo de compromiso, no obstante, se va a perseguir fundamentalmente a través de dos mecanismos que, precisamente por ello, serán los más atacados y cuestionados en las acciones judiciales. Se trata, en primer lugar, de la creación y mantenimiento de flujos de noticias o contenidos adictivos a través de sofisticados sistemas de recomendación algorítmica y, en segundo lugar, del diseño consciente o intencionado de las plataformas para favorecer el uso continuado, e incluso compulsivo, mediante recursos técnicos, como el diseño de las interfaces o los scrolls infinitos y recursos psicológicos¹². En este sentido, algunos de los mecanismos o características de las redes que refuerzan la atención y la permanencia del usuario son: las recompensas impredecibles, generalmente basadas en técnicas de reafirmación y validación social¹³; el miedo a estar ausente o perderse algún contenido o información relevante (conocido en inglés por las siglas *fomo* o “*fear of missing out*”); la conocida, en una traducción literal del inglés, como gamificación, técnica consistente en añadir juegos o elementos propios de los juegos a una tarea o actividad para fomentar la implicación (p.e. ruleta de premios de Temu o el sistema de recompensas denunciado por la Comisión europea, TikTok Lite), los sonidos y vibraciones de los dispositivos, likes recíprocos, competencia social e inversión psicológica (por ejemplo, los sistemas de rachas de Snapchat o Duolingo).

23. El segundo pilar del modelo empresarial de estas plataformas consiste en la monetización de la atención, que se lleva a cabo, principalmente mediante publicidad, sin perjuicio de que las plataformas se puedan financiar de otras maneras, por ejemplo, a través de suscripciones como es el caso de Netflix, modelos de acceso escalonado o mediante la creación de bolsas de contenido exclusivo.

24. Al modelo descrito pertenecen las redes sociales Facebook, Instagram, YouTube y TikTok, que si bien, no son las únicas, sí han sido las principales destinatarias de las acciones judiciales.

2. Las principales demandadas: Meta (Facebook e Instagram), Alphabet (YouTube) y Bytedance (TikTok)

25. La red social Facebook fue fundada en 2004 por Mark Zuckerberg y otros compañeros universitarios, en principio, como una red privada de intercambio de información entre estudiantes de Harvard. Debido a su éxito, en pocos meses se amplió su uso a otras universidades en el área de Boston, hasta volverse popular en todas las universidades estadounidenses. En septiembre de 2006 se permitió el acceso del público general al proyecto hasta convertirse, hoy en día, en la red social con el mayor número de usuarios registrados del planeta, con **más de 2.800 millones de usuarios activos mensuales**. En 2012 Facebook adquiere Whatsapp e Instagram en 2021. Ese mismo año decide realizar cambios en su marca y abandona el nombre de Facebook, manteniéndolo solo para la red social y adopta el de Meta para recalcar la planeada expansión de la empresa en la realidad virtual y el metaverso.

26. Tanto Facebook como Instagram son plataformas de la atención, en el sentido de que están orientadas a conseguir el compromiso del usuario que es monetizado fundamentalmente mediante publicidad personalizada.

¹² Vid. M.D. GRIFFITHS: “Weapons of Mass Distraction and Addiction by Design: What Techniques do Social Media Operators use to increase User Engagement?”, *TechReg Chronicle*, 2024.

¹³ Las cualidades adictivas de las plataformas se basan en gran medida en sus programas de recompensa impredecibles, centradas, en gran medida, en la validación social. Las notificaciones, los “likes”, los “retweets”, los comentarios y otros elementos del diseño digital, de aparición variable, proporcionan gratificaciones inciertas, lo que lleva a los usuarios a revisar repetidamente sus redes sociales en busca de esa sensación de satisfacción. Vid. A., ALTER: *Irresistible. ¿Quién nos ha convertido en yonquis tecnológicos?*, Paidós, Barcelona, 2017, pp. 117 y ss.

27. YouTube es una plataforma a través de la que reproducir vídeos, subir contenido a un canal propio e interactuar con el resto de usuarios. Creada en 2005 por tres ex empleados de PayPal es adquirida un año después por Google. Su modelo de negocio es generar atención y participación en torno al intercambio y visualización de contenido generado por usuarios y socios, monetizándolas, principalmente a través de anuncios¹⁴ y, en menor medida, de suscripciones como en el caso del servicio YouTube Premium que ofrece streaming sin publicidad en todos los vídeos. La plataforma comenzó su actividad como un sitio para compartir contenido generado por el usuario y amplió progresivamente su modelo de negocio basado en publicidad y sus capacidades de contenido exclusivo y pago con el programa de socios.

28. Tik Tok es una red social, propiedad del gigante tecnológico chino ByteDance. La plataforma se utiliza para hacer videos de formato corto y vertical. La plataforma nace en 2016 como una red social centrada en la educación con la finalidad de compartir vídeos cortos para que los usuarios enseñaran o aprendieran de distintas temáticas. No consiguió el favor del público y se optó por un nuevo enfoque, mezclando vídeos y música en la aplicación. Su fusión en 2018 con otra red social china, Musical.ly, la catapultó al mercado internacional con el nombre de TikTok. Desde entonces, ha ido creciendo de forma exponencial hasta alcanzar en 2024, la cifra de 1.218 millones de usuarios activos al mes en todo el mundo. TikTok se financia también principalmente a través de anuncios¹⁵. Como hemos señalado, la batalla judicial contra las redes sociales no es exclusiva de Tik Tok, pero en esta plataforma confluyen dos características que condicionan el alcance de estas e introducen nuevos parámetros en la valoración de sus conductas y en las causas de actuación: 1) la nacionalidad china de la plataforma y 2) su difusión y éxito entre los menores de edad y, muy especialmente, entre los menores de 13 o 14 años. En este sentido, el temor a un posible acceso y utilización por el gobierno chino de datos personales e información de los usuarios recabados por la plataforma ha determinado que numerosos gobiernos y administraciones públicas hayan prohibido a sus empleados el uso de la plataforma en dispositivos profesionales o incluso personales (con aplicaciones profesionales). Este es el caso de la Comisión Europea, el Parlamento y el Consejo, los gobiernos de Bélgica, Dinamarca, Canadá, Reino Unido, Australia, Nueva Zelanda, Noruega, Taiwan y EE.UU. Algunos países han dado un paso más allá y han prohibido directamente el uso de la aplicación en sus territorios, es el caso de India, Pakistan, Nepal, Somalia y muy significativamente EE.UU. con una ley aprobada en mayo de 2024 que obligaba a la empresa matriz a vender la aplicación en un plazo de nueve meses o de lo contrario, la app sería prohibida en el país. Ante el incumplimiento de la norma, ratificada por unanimidad por el Tribunal Supremo, Tik Tok fue suspendida en EE.UU., pero solo durante trece horas, dado que el ejecutivo entrante de Donald Trump decidió desactivar, por el momento, la prohibición.

29. En septiembre de 2023, la Comisión Europea incluyó a Bytedance como guardián de acceso o *gate keeper* con respecto a Tik Tok (servicio básico de plataforma), a Meta con, entre otros, Instagram y Facebook y a Alphabet, matriz de Google, con YouTube, en aplicación del Reglamento de Mercados Digitales¹⁶. Igualmente, todas ellas han sido designadas como *Plataformas Online de Gran Tamaño* (VLOP) con arreglo al Reglamento de Servicios Digitales¹⁷, por lo que todo el paquete regulatorio digital europeo es de aplicación a estas plataformas.

¹⁴ Es posible anunciarse en YouTube de varias formas: dentro o antes de un vídeo (saltable o no saltable), a través de un banner, cerca o en la pantalla del vídeo, o en la página de búsqueda junto a los resultados.

¹⁵ La publicidad no es, no obstante, la única forma de financiación de TikTok, que también cuenta con otras fuentes de ingresos, como la oferta de compras dentro de la aplicación y de recompensas virtuales para la participación del usuario. La plataforma, además, colabora con marcas para desafíos de contenido patrocinado y vende moneda virtual (TikTok coins) para dar propinas a los creadores de contenido.

¹⁶ Reglamento (UE) 2022/1925 del Parlamento Europeo y del Consejo de 14 de septiembre de 2022 *sobre mercados disputables y equitativos en el sector digital y por el que se modifican las Directivas (UE) 2019/1937 y (UE) 2020/1828* (Reglamento de Mercados Digitales).

¹⁷ Reglamento (UE) 2022/2065 *relativo a un mercado único de servicios digitales y por el que se modifica la Directiva 2000/31/CE* (Reglamento de Servicios Digitales)

III. Acciones judiciales contra las redes sociales

1. De la privacidad a la adicción

30. Inicialmente, los principales temores frente a las redes sociales se centraron, tanto judicial como normativamente, en el aspecto más vinculado a la economía de la vigilancia: el tratamiento dado a los datos personales de sus usuarios. La falta de seguridad, la violación del derecho a la privacidad y el uso que las plataformas habían hecho de los datos personales ha sido el fundamento de diversas demandas contra las redes sociales en la última década.

1. De la privacidad a la adicción

31. En este sentido, si tomamos como punto de referencia la red social china TikTok, contra esta plataforma se han entablado diversos procedimientos, tanto administrativos como civiles resarcitorios, en los últimos cinco años, por vulneración de las normas en materia de protección de datos. En EE.UU., Tik Tok fue sancionada en 2019 por la FTC con 5,7 millones de dólares por recopilar información personal de menores en violación de la *Ley de protección de la privacidad en línea de los niños* (COPPA, por sus siglas en inglés). Solo cinco años después, en agosto de 2024, el DOJ, en representación de la FTC ha demandado a la compañía, por infringir nuevamente la COPPA y la orden de 2019 de la FTC. Pero no solo hay procedimientos administrativos, en 2021, se produjo un acuerdo transaccional (*settlement*) de 92 millones de dólares en el marco de una acción colectiva contra la empresa china por la utilización ilícita de datos privados e información biométrica de sus usuarios mediante su cesión a terceros¹⁸. En Europa destaca la multa de 345 millones de euros impuesta en 2023 por la UE a través de la Agencia de Protección de Datos irlandesa por vulnerar el derecho a la intimidad de los menores en infracción del Reglamento Europeo de Protección de Datos¹⁹. La Comisión Europea ha abierto un procedimiento contra TikTok, en aplicación de la Ley de Servicios Digitales, para examinar si la plataforma ha incumplido la obligación de garantizar un alto nivel de privacidad y protección del menor, en especial, en relación a la configuración de la privacidad por defecto²⁰.

32. Pero la lesión a la privacidad no es el único argumento de ataque a las plataformas contra las que se han entablado procedimientos, tanto de naturaleza administrativa como resarcitorios, en base a distintos argumentos, centrados en **los posibles daños que pueden derivarse de su uso para la salud mental y física de sus usuarios**, especialmente en el caso de los menores de edad. Estas demandas, resultan, a mi juicio, más retadoras desde una perspectiva jurídica, obligando a buscar el encuadre de estas pretendidas lesiones en nuestro ordenamiento jurídico, la posibilidad de exigir o no responsabilidad a la plataforma y, en su caso, por qué vías y con qué límites.

33. En **Europa** existen diversos casos, con un incremento progresivo. En este sentido, tanto en Portugal²¹ como en los Países Bajos²² se han entablado acciones colectivas contra la red social Tik Tok, de carácter mixto, dado que recurren a consideraciones tanto de competencia desleal como de protección de datos para alegar daños a la privacidad de los menores y a su salud mental. Desde una perspectiva administrativa, en Italia, la Autorità Garante della Concorrenza impuso una multa de 10 millones de euros a TikTok, por los daños psico-físicos que sufrieron menores de edad por participar en un reto viral, conocido como la “cicatriz francesa”, consistente en lesionarse los pómulos, dejando marcas rojizas en el

¹⁸ In re TikTok, Inc., 565 F. Supp. 3d 1076

¹⁹ Decisión vinculante 2/2023, de 2 de agosto de 2023, *sobre la controversia planteada por la AC irlandesa en relación con Tik Tok Technology Limited* (art.65 RGPD).

²⁰ Vid. https://ec.europa.eu/commission/presscorner/detail/es/ip_24_926.

²¹ <https://tiktok.iusomnibus.eu>.

²² <https://www.massaschadeconsument.nl/collectieve-acties/tiktok/>

rostro²³. Por su parte, la Comisión Europea entre los puntos objeto de investigación formal para evaluar si TikTok pudo haber infringido la Ley de Servicios Digitales, incluyó la valoración de si TikTok había evaluado y mitigado adecuadamente los riesgos sistémicos asociados con el diseño de un nuevo servicio TikTok Lite, incluyendo los sistemas algorítmicos que pueden estimular adicciones conductuales. De hecho, Thierry Breton, Comisario de Mercado Interior, ha afirmado expresamente: “Consideramos que TikTok Lite puede ser tan tóxico y adictivo como los cigarrillos *light*”. También Meta está siendo investigado por la Comisión para determinar si ha incumplido la LSD en relación con la protección de los menores en Facebook e Instagram. A la Comisión le preocupa que los sistemas y algoritmos de Facebook e Instagram puedan estimular adicciones en el comportamiento de los niños, así como crear los denominados “efectos agujero de conejo” (*rabbit hole effects*)²⁴. Se investigan además los métodos de garantía y verificación de la edad establecidos por Meta²⁵.

2. Demandas sobre adicción a las redes sociales en EE. UU. (*social media addiction lawsuits*).

34. No obstante, y pese al progresivo incremento de las demandas en Europa en los últimos años, es sin duda EE. UU., quien más procedimientos de daños tiene entablados contra las redes sociales, tanto individual como conjuntamente como codemandadas. Nos encontramos con una avalancha de demandas, a nivel federal y estatal (California, Nueva York, Seattle, Nuevo México, etc.) interpuestas por particulares, centros educativos, ayuntamientos y Fiscales Generales, en aplicación del derecho del consumo, competencia desleal y responsabilidad extracontractual (*torts*): general o basado en responsabilidad derivada de producto defectuoso (*products' liability*) e, incluso, por alteración del orden público (*public nuisance*), lo que puede ser perseguido tanto en vía penal como civil.

35. En estas demandas se exige a las plataformas modificaciones en sus modelos de negocios y compensación por los daños sufridos por los ciudadanos estadounidenses, especialmente adolescentes y menores de edad, en su salud mental y física, derivados del desarrollo de trastornos como ansiedad, depresión, comportamientos compulsivos, autolesiones o incluso suicidio en los casos más graves, como consecuencia del diseño adictivo de las plataformas. Estas demandas se conocen, en la jerga periodística y de los despachos de abogados, como *social media addiction lawsuits* o demandas sobre adicción a las redes sociales.

36. Frente a estas demandas, las plataformas han mantenido una primera línea de defensa uniforme: la ausencia de responsabilidad en aplicación de la **enmienda constitucional sobre libertad de expresión**²⁶ y, principalmente, del **artículo 230 de la *Communication Decency Act*** de 1998²⁷.

²³ Provvedimento n. 31124 del 5 de marzo de 2024, PS12543 - TIKTOK CICATRICE FRANCESE.

²⁴ El “rabbit hole effect” o efecto agujero de conejo en las redes sociales se refiere a la tendencia de los usuarios a caer en un ciclo de consumo de contenido que los lleva a explorar cada vez más temas relacionados, a menudo de manera profunda y sin conciencia del tiempo empleado.

²⁵ Vid. https://ec.europa.eu/commission/presscorner/detail/es/ip_24_2664.

²⁶ Los intentos de regular el daño derivado de las plataformas de redes sociales corren el riesgo de violar los derechos de libertad de expresión de la Primera Enmienda debido a las restricciones que buscan imponer al contenido, específicamente a los algoritmos, utilizados por las plataformas de redes sociales. Sobre el alcance del derecho de libertad de expresión en este contexto, Vid. in extenso, COSTELLO ET ALTRI: “Algorithms, Addiction and Adolescent Mental Health: An Interdisciplinary Study to Inform State-level Policy Action to Protect Youth from the Dangers of Social Media”, *American Journal of Law and Medicine*, 49, 2023, pp. 148 y ss.; M.B. LAWRENCE: “Public Health Law’s Digital Frontier: Addictive Design, Section 230 and the Freedom of Speech”, disponible en SSRN: <https://ssrn.com/abstract=4603784>, pp. 29 y ss.

²⁷ El recurso a la libertad de expresión o a la sección 230 de la CDA son un primer intento de frenar la acción de responsabilidad que impiden, de tener éxito, entrar siquiera a valorar los elementos de la pretensión resarcitoria. Para el caso de que estas defensas no sean operativas, los demandados también cuestionan habitualmente, la inexistencia de un producto al que sean de aplicación las normas de responsabilidad por producto defectuoso, de un nexo causal adecuado y suficiente o de un deber de conducta.

37. Este artículo proporciona inmunidad a los servicios informáticos en línea con respecto a los contenidos generados por terceros usuarios, de forma similar a los puertos seguros recogidos en los artículos del Reglamento europeo de servicios digitales²⁸. Su aplicabilidad a las redes sociales ha sido establecida judicialmente en numerosos pronunciamientos²⁹.

38. Si bien aún nos encontramos en los estadios iniciales en gran parte de estas demandas, algunas resoluciones recaídas en el marco de acciones individuales, al igual que diversos pronunciamientos judiciales en los grandes casos contra las redes, valorando las peticiones de desestimación de la demanda, especialmente en base al mencionado artículo 230 CDA, posibilitan apuntar ya algunas reflexiones sobre la viabilidad de estas demandas y su posible traslado al marco europeo.

39. Para ello, es importante distinguir los supuestos que, en ocasiones, se entremezclan en las distintas demandas entabladas contra las redes sociales.

A) Daños derivados del contenido

40. En primer lugar, algunas de las reclamaciones se refieren a daños directamente derivados de la exposición de los usuarios (en la mayoría de los casos menores de edad) a determinados contenidos. Este es el escenario, por ejemplo, en:

²⁸ A finales de la década de 1990, comenzaron a plantearse en los tribunales estadounidenses una serie de casos relacionados con la posible responsabilidad de los intermediarios de Internet por contenidos difamatorios transmitidos a través de la red. Los jueces intentaron resolverlos aplicando los principios del *common law* para la difamación que, a pesar de basarse en el principio de que la persona que repite o de cualquier otra forma reproduce o publica material difamatorio tiene la misma responsabilidad que el autor de este material, diferencia tres niveles de responsabilidad. En primer lugar, la responsabilidad del editor, que al tener control sobre los contenidos tiene la responsabilidad principal; la responsabilidad del distribuidor, que solo es responsable si tiene conocimiento de la difamación o tiene alguna razón para tener conocimiento de la misma, y la responsabilidad del mero canal de alojamiento (*common carrier*), que no se hace responsable independientemente de la existencia o no de conocimiento de la difamación. En un caso temprano, *Cubby contra CompuServe* (776 F.Supp. 135, S.D.N.Y. 29 de octubre de 1991), el tribunal concedió una importante salvaguardia a los intermediarios de Internet, al no considerarlos editores, sino meros distribuidores de información de terceros que, por lo tanto, estaban sujetos a un régimen de responsabilidad mucho más clemente. Años más tarde, esta doctrina se vió gravemente afectada, como resultado de un caso controvertido, el caso *Stratton Oakmont vs. Prodigy*, 1995 WL 323710 (N.Y. Sup. Ct. 1995) que identificó a un operador de telecomunicaciones como editor y, por lo tanto, le atribuyó la misma responsabilidad por una acción de difamación que a su autor material. El criterio del tribunal fue que, en este caso, el operador tenía un cierto grado de control sobre los contenidos, como resultado de haber definido una serie de mecanismos para detectar y eliminar contenidos ilícitos (software específico, un moderador del foro, etc.), y que, como resultado, debía ser considerado como un editor. Este caso abrió la puerta a una situación paradójica: el operador de Internet que no hubiera ejercido ningún tipo de control sobre los contenidos probablemente sería considerado como un distribuidor y podría eludir su responsabilidad, mientras que, por el contrario, el operador que hubiera intentado evitar la publicación de contenidos ilícitos tendría que responder por los contenidos cuya publicación no hubiera podido impedir.

Como reacción a esta situación resultante del caso *Stratton*, en 1996 se aprobó la *Communication Decency Act*, que añadía una nueva sección (230) a la *Communication Act* de 1934, definiendo la responsabilidad de los proveedores de servicios de la sociedad de la información por contenidos de terceros, desde una doble perspectiva. En primer lugar, se estableció que el proveedor de un servicio informatizado no sería considerado como el editor o emisor de cualquier información suministrada por otro proveedor de contenido informativo. Una segunda norma, conocida como «protección del buen samaritano», elimina la doctrina *Stratton* al determinar que el proveedor de un servicio informático no puede ser considerado responsable de restringir de buena fe el acceso o la disponibilidad de material censurable, o de suministrar los medios técnicos para restringir el acceso a este material. Aquí el objetivo era fomentar la participación de los proveedores de servicios en la detección de contenidos ilícitos, precisamente con el fin de garantizar que este control editorial de los contenidos pudiera dar lugar a una acusación de responsabilidad.

Estas exclusiones de responsabilidad se han interpretado de manera amplia, lo que ha favorecido en gran medida la posición de los proveedores de servicios de intermediación. En este caso, la posición dominante en la jurisprudencia, basada en una interpretación amplia del término «editor», ha sido considerar que ningún intermediario puede ser considerado responsable de los contenidos proporcionados por un tercero, ni en base a la responsabilidad estándar del editor, ni aplicando la responsabilidad estándar del distribuidor común. Su postura se compara con la de un mero alojador o «transportista» (*common carrier*) que no sería considerado responsable incluso si supiera o tuviera alguna razón para saber de la naturaleza ilegal de los datos contenidos en sus redes.

²⁹ Vid. *ad.ex.*, *Klayman v. Zuckerberg*, 753 F.3d 1354, 1355 (D.C. Cir. 2014) (considerando a Facebook como un servicio informático en línea); *Fields v. Twitter, Inc.*, 217 F. Supp. 3d 1116, 1118 (N.D. Cal. 2016) (considerando a Twitter como un servicio informático en línea).

- 1) Los mediáticos casos de los retos virales o «challenges» que consisten en desafíos o pruebas que se graban con un dispositivo digital y se cuelga en alguna red social, invitando a otros usuarios a realizar la misma conducta. Muchos de estos juegos popularizados en redes implican graves riesgos para la salud, pudiendo causar lesiones o, incluso la muerte. Los ejemplos son muy numerosos, solo por citar algunos de los más populares: *chroming*: el desafío consiste en inhalar productos químicos tóxicos, como aerosoles de pintura, esmaltes o desodorantes; “*el que duerma el último gana*”: Lo que sugiere el desafío es tomar Clozapepam o Diazepam en grupo y el que consiga ver como cae el resto, y por tanto aguante más sin dormirse, gana., o el muy popular “Abecedario del diablo”, que anima a cortarse con algo punzante mientras se recita el abecedario y no parar hasta terminar de decirlo.
- 2) Exposición a páginas o contenidos relacionados con desórdenes alimenticios, (dietas extremas, técnicas para vomitar, body shaming, imágenes demacradas³⁰) o de tendencia suicida. Por ejemplo, en enero de 2022, Tammy Rodriguez demandó a Meta, propietaria de Instagram, y Span Inc., a la que pertenece Snapchat, acusándolas de propiciar el suicidio de su hija Selena, de solamente once años³¹. También contra Meta es la demanda, interpuesta ante el Tribunal del Distrito Norte de California, por la asociación Social Media Victims Law Center, en representación de Alexis Spence y sus padres, por los daños sufridos por la niña en su salud física y mental como consecuencia de la exposición a determinados contenidos en Instagram. En la demanda se alega que al poco de abrirse la cuenta, casi de forma inmediata el motor de inteligencia de Instagram condujo a Spence, entonces de once años de edad, a grupos con contenidos que glorificaban la anorexia o las autolesiones, a la vez que fomentaba una adicción de la niña a la aplicación³².
- 3) Exposición a contenidos falsos, extremos o manipulados. Esta es la situación considerada, por ejemplo, en la denuncia de Amnistía Internacional en el caso *Rohinyá*. El conflicto rohinyá es un conflicto en curso en la parte norte de Myanmar, caracterizado por la violencia sectaria entre las comunidades musulmana rohinyá y budista rakhine y la represión militar de los civiles rohinyá por parte de las fuerzas de seguridad de Myanmar. El conflicto ha provocado la migración masiva de rohinyás desde Myanmar, a partir de 2015, y varias masacres de rohinyás por parte del ejército de Myanmar y de lugareños armados en 2017. En 2022, Amnistía Internacional acusó a Meta, la empresa matriz de Facebook, de haber «contribuido sustancialmente» a las violaciones de los derechos humanos perpetradas contra el pueblo rohinyá de Myanmar. Amnistía afirmó que los algoritmos de Facebook «amplificaron proactivamente» el contenido contra los rohinyás. También alegó que Meta ignoró las peticiones de civiles y activistas para frenar la incitación al odio en la plataforma de redes sociales mientras se beneficiaba del aumento de compromiso o *engagement* de los usuarios³³.

41. En esta primera categoría de acciones, especialmente en los dos primeros supuestos o grupos, si bien puede ser más fácil apreciar el nexo de causalidad entre el daño y la exposición a un determinado contenido, también hay más probabilidad de que operen los puertos seguros. De hecho, los tribunales en EE.UU. han desestimado muchas demandas contra las redes sociales (p.e. *challenges*) en aplicación de la sección 230 CDA y la no responsabilidad de la plataforma por los contenidos creados por sus usuarios. No obstante, el alcance de este escudo está siendo objeto de cuestionamiento en EE.UU., debatiéndose si la inmunidad que otorga este artículo también operaría si la plataforma no se limita simplemente a poner a disposición del usuario contenido creado por terceros (o a no retirar contenido de terceros) sino que realiza, además, una labor de recomendación o “matchmaking” o “emparejamiento” conectando a

³⁰ Vid. La investigación realizada por el *Wall Street Journal* el 17 de diciembre de 2021, “The Corpse Bride Diet: How TikTok Inundates Teens with Eating-Disorder Videos”, disponible en <https://www.wsj.com/articles/how-tiktok-inundates-teens-with-eating-disorder-videos-11639754848>.

³¹ <https://www.courthousenews.com/wp-content/uploads/2022/01/rodriguez-meta-snap-complaint.pdf>

³² https://socialmediavictims.org/wp-content/uploads/2022/06/Spence-Complaint-6_6_22.pdf

³³ Vid. <https://www.amnesty.org/es/latest/news/2022/09/myanmar-facebooks-systems-promoted-violence-against-rohiny-gya-meta-owes-reparations-new-report/>

los usuarios con un determinado contenido, basándose en la probabilidad de que, en función de los datos de los que dispone y del algoritmo, este le interese³⁴.

42. El Tribunal Supremo estadounidense cerró sin pronunciarse un caso en el que, precisamente, la cuestión examinada era si las plataformas podían ser consideradas responsables, es decir, no estar protegidas por la inmunidad de la sección 230 CDA, cuando promueven de forma activa contenido ilícito mediante recomendaciones algorítmicas. Se trata del mediático caso *González v. Google*³⁵. En noviembre de 2015, una serie de atentados del Estado Islámico (ISIS) mataron a más de 130 personas en París, entre ellas la ciudadana estadounidense Nohemi González, una estudiante universitaria californiana de 23 años que estaba en un programa de intercambio. La familia de González demandó a Google en virtud de una disposición de la ley federal que impone responsabilidad a cualquiera que ayude, intencionadamente o no, a un ataque terrorista. La familia argumentó que Google era responsable porque promocionaba contenido de ISIS en su plataforma de YouTube a los usuarios mediante algoritmos de recomendación, ayudando así a los esfuerzos de reclutamiento de ISIS y, por lo tanto, facilitando, aunque fuera de manera indirecta, a ISIS la realización de los ataques del 2015. El caso causó una gran expectación, una eventual interpretación restrictiva de la CDA podría revolucionar la red, exigiendo a las plataformas una reformulación de sus modelos de negocios. Después de una compleja vista en la que los propios magistrados reconocieron tanto la situación de especial privilegio de los operadores tecnológicos que, gracias al amplísimo paraguas de la sección 230 CDA, no internalizaban sus costes, como la complejidad de su supresión o limitación en la que estaban en juego libertades fundamentales, el Tribunal Supremo decidió “no decidir”. El 18 de mayo de 2023, en una breve opinión consideró que, dado que la causa y petición de los demandantes, aunque no fuera de aplicación la sección 230, no tenía posibilidades de prosperar, las cuestiones respecto al alcance y aplicabilidad de la sección planteadas en el caso no estaban listas para una decisión³⁶.

43. Por tanto, en ausencia de un pronunciamiento expreso del Tribunal Supremo, la cuestión sigue abierta. Si bien no hay una línea uniforme de razonamiento, existiendo posturas discrepantes en los tribunales de Circuito³⁷, al menos, comienza a resquebrajarse la política de aplicabilidad prácticamente automática de la sección 230 CDA, a los algoritmos personalizados. En este sentido, es interesante la reciente sentencia del Tribunal de Apelaciones del Tercer Circuito en el caso *Tawainna Anderson v. TikTok*³⁸. TikTok, a través de su algoritmo, recomendó en su página personalizada (*For You Page*) a Nylah Anderson, una niña de diez años, un video con el reto del *blackout* o desmayo que anima a los espectadores a realizar actos de auto asfixia hasta provocar un desmayo. La niña murió ahorcada y su madre demandó a TikTok y a su empresa matriz Bytedance por responsabilidad extracontractual (*strict products liability* y *negligence*). En primera instancia la demanda fue desestimada en aplicación de la sección 230 CDA, al considerarse que esta excluía e inmunizaba a la empresa de cualquier tipo de res-

³⁴ Vid., también en la doctrina más reciente, posiciones discrepantes, T. KENNETH/I. RUBINSTEIN: “Gonzalez v. Google: The Case for Protecting “Targeted Recommendations”, *Duke Law Journal Online*, vol.72, 2023, pp. 177 y ss., favorables a la aplicabilidad del puerto seguro de la CDA a los algoritmos personalizados o recomendados, frente a posiciones más restrictivas, A.Z.ROZENSHTAIN; “Interpreting the Ambiguities of Section 230”, *Yale Journal on Regulation Bulletin*, núm.41, 2024, pp. 60 y ss.; A. ZAKON: “Optimized for Addiction: Extending Product Liability Concepts to Deffectively Designed Social Media Algorithms and Overcoming the Communications Decency Act”, disponible en SSRN: <https://ssrn.com/abstract=3682048>.

³⁵ *Gonzalez v. Google, LLC*, 598 U.S. 617 (2023).

³⁶ «Por lo tanto, nos negamos a abordar la aplicación de la sección 230 a una demanda que parece presentar pocas, o ninguna, reclamación plausible de reparación».

³⁷ Vid. *ad.ex. Force v. Facebook, Inc.*, 934 F.3d 53 (2d Cir 2019), donde se rechaza la inaplicabilidad del 230CDA incluso en los casos en los que la exposición a un determinado contenido no es el resultado de una búsqueda activa del usuario, sino una recomendación de la plataforma, considerándose que también en este último supuesto, la plataforma se limita a difundir contenido de terceros. La creación de algoritmos que ofrezcan, categoricen y promuevan contenido dañino probablemente no se consideraría un acto de provisión de contenido, sino más bien una función editorial protegida de responsabilidad, incluso si una plataforma de redes sociales obtiene ganancias de los ingresos publicitarios. Las plataformas pueden estar ayudando y beneficiándose de la orientación de contenido dañino a usuarios menores, pero no están creando el contenido en sí y, por lo tanto, pueden estar inmunizadas bajo la sección 230

³⁸ *Tawainna Anderson v. TikTok Inc, Bytedance, Inc.*No. 22-3061 (3d Cir. 2024).

ponsabilidad. No obstante, en apelación, el tribunal de Circuito rechaza la aplicabilidad en este supuesto del puerto seguro, en una reinterpretación del alcance y función de esta cláusula. Es interesante esta sentencia, porque dos son las teorías adelantadas para fundamentar su inaplicación, una mayoritaria y otra recogida en el voto particular de un magistrado.

44. La mayoritaria se basa en un pronunciamiento del Tribunal Supremo en *Moody v. NetChoice, LLC*³⁹, en el que, en un caso sobre libertad de expresión, la Corte examina la naturaleza de los algoritmos de redes sociales que construyen flujos de información (o *feeds*) para transmitir y recomendar contenido a los usuarios. La labor selectiva del contenido de terceros y la decisión sobre su presentación (que implican juicios editoriales) determina que este resultado pueda ser considerado “discurso” “contenido” o “expresión” de la plataforma (por tanto, en este supuesto, discurso protegido por la primera enmienda que reconoce el derecho a la libertad de expresión). El Tribunal de Circuito, siguiendo este razonamiento, defiende la inaplicabilidad de la section 230 CDA, dado que esta solo opera frente a contenido de terceros, pero no frente a contenido de la propia plataforma como sería el algoritmo personalizado.

45. Resulta más interesante, a mi juicio, el voto particular del juez MATEY, que realiza una interpretación teleológica y sistemática del precepto, rechazando la aplicación automática y acrítica que están realizando numerosos jueces frente a cualquier demanda a las redes sociales.

46. *“Hoy en día, la sección 230 viene al rescate de las empresas de prácticamente cualquier reclamación vagamente relacionada con el contenido publicado por un tercero, sin importar la causa de la acción y cualesquiera que sean las acciones del proveedor. El resultado es una sección 230 que inmuniza a las plataformas de las consecuencias de su propia conducta y les permite ignorar la obligación ordinaria que la mayoría de las empresas tienen de tomar medidas razonables para evitar que sus servicios causen daños devastadores”*⁴⁰.

47. En una interpretación restrictiva de la sección, considera que esta solo protege al proveedor cuando se limita a alojar contenido de terceros. Por tanto, Tik Tok no puede ser considerado responsable por la mera presencia del video del reto de Blackout en su plataforma, pero esto no significa que no pueda tener responsabilidad como “distribuidor” por mantener esos videos en la plataforma y, sobre todo, recomendarlos siendo conscientes del daño que podían producir. Esto no es una responsabilidad por conducta de terceros sino por la propia conducta de la plataforma. Si esta es ilícita o no, es otra cuestión...lo que se debate es la operatividad o no de ese primer muro que impide entrar siquiera a analizar la conducta.

B) Daños derivados del carácter adictivo de la plataforma

48. Dadas las dificultades expuestas, para intentar eludir los muros de la Primera enmienda y, sobre todo, de la sección 230 CDA, en un segundo escalón de demandas se van a buscar otras bases de imputación de responsabilidad no relacionadas con el contenido difundido en la plataforma. Así, un segundo parámetro de acusación se basa en los daños asociados al carácter adictivo derivado de la propia configuración de la tecnología de la plataforma⁴¹. Dado que los beneficios de las plataformas

³⁹ *Moody v. NetChoice, LLC and NetChoice, LLC v. Paxton*, 603 U.S. 707 (2024).

⁴⁰ *“Today, section 230 rides in to rescue corporations from virtually any claim loosely related to content posted by a third party, no matter the cause of action and whatever the provider’s actions. The result is a section 230 that immunizes platforms from the consequences of their own conduct and permits platform to ignore the ordinary obligation that most business have to take reasonable steps to prevent their services for causing devastating harm”*.

⁴¹ Las bases de imputación son muy variadas. En algunos casos, se acude a las normas de Derecho privado en materia de responsabilidad extracontractual, ya sea las normas de responsabilidad por los daños causados por productos defectuosos, las normas sobre daños extracontractuales o torts con carácter más general (*negligence theory*) o incluso, algunos demandantes han acudido a las normas que sancionan las alteraciones o lesiones del orden público (*public nuisance*), concretamente esta ha sido la base elegida por numerosos centros educativos, la ciudad de Nueva York contra una pluralidad de redes sociales o, por ejemplo, el Estado de Massachusetts contra Meta.

están directamente relacionados con el consumo y tiempo que los usuarios pasan en ella, las empresas de social media están continuamente adaptando su tecnología para maximizar el *engagement* o nivel de compromiso del usuario, diseñando conscientemente el producto para promover su uso compulsivo mediante la alteración de sus niveles de dopamina de tal modo que los usuarios pueden experimentar cambios en las estructuras cerebrales similares a los experimentados en el consumo de diversas sustancias o adicciones al juego.

49. Las afirmaciones de diseño adictivo alegan que las plataformas participan en lo que los psicólogos llaman condicionamiento operante mediante el uso de refuerzo intermitente de contenido neutral y técnicas de recompensa variable (similar a lo que se utiliza, por ejemplo, en relación con las máquinas tragaperras), pero, en estos casos con mecanismos como *scroll* infinito o *likes*, para fomentar la compulsión en los usuarios. Características como el desplazamiento infinito, el feed de noticias, las funciones de ‘me gusta’ y reposteo, el hecho de que cierta información solo esté disponible temporalmente, o que sea visible para una gran multitud, o que uno reciba información sobre la recepción y visualización de otros usuarios (doble tilde azul) e interacción con el contenido (respuestas y notificaciones) tienden a producir un compromiso continuo. Estas técnicas no tienen por qué implicar moderación de contenido o expresión editorial; de hecho, normalmente están ocultas a los usuarios, quienes tal vez nunca se den cuenta de que han sido condicionadas por un proveedor. Los demandantes alegan que es la propia adicción a la red la que tiene efectos nocivos para la salud mental y física de los usuarios, pudiendo dar lugar, entre otros, a trastornos como ansiedad, desórdenes alimenticios, depresión, autolesiones o, incluso, en los casos más graves, al suicidio.

50. Se busca un paralelismo entre el uso adictivo a las redes sociales por parte de los menores con la adicción al juego u a otras drogas denominadas “sin sustancia” (por ejemplo, la ludopatía), con efectos psicológica y neurológicamente análogos. En estos casos, no se cuestiona directamente el específico contenido presentado o el grado en que una plataforma controla el contenido, censurando o evitando censurar, priorizando o recomendando un contenido particular.

51. ¿Cuál es la operatividad de la section 230 CDA (y, en su caso de la Primera Enmienda) en estos casos? ¿Puede funcionar igualmente como escudo protector de las plataformas garantizándoles inmunidad frente a estas acusaciones? Se está exigiendo a la plataforma responsabilidad, no por su relación con el contenido de terceros, sino por las características y funciones de sus propias aplicaciones.

52. Por el momento, la respuesta de las autoridades judiciales no está siendo concluyente a este respecto, fundamentalmente encontramos resoluciones valorando peticiones de desestimación de las demandas de diseño adictivo, en las que, si bien hay una cierta apertura en la interpretación de la section 230 CDA, esta resiste alguno de los embates, fundamentalmente derivados de la dificultad -alegada por las plataformas- de determinar que rasgos o características de las plataformas pueden deslindarse del “contenido de terceros” y de la función como mero editor. Destacan, a este respecto, la resolución del Tribunal Superior del Condado de Suffolk y las resoluciones del Tribunal de Distrito Norte de California, en las que se mantiene una posición ambigua, pero se rechaza la operatividad automática frente a estas demandas de la section 230 CDA. Con arreglo al Tribunal de California hay rasgos denunciados como causantes de la adicción que involucran necesariamente un proceso relacionado con la publicación y, por tanto, están exentos: *scroll* infinito, contenido efímero, las notificaciones y los likes. Otros, por el contrario, no están relacionados con la publicación y, por tanto, protegidos por la inmunidad, como la insuficiencia de controles parentales robustos y el momento y agrupamiento de las notificaciones del contenido propio de los demandados⁴².

53. Pese a que no se refiere directamente al riesgo de que el diseño adictivo de las redes sociales cause daños, es interesante la sentencia del Tribunal del Noveno Circuito de 4 de mayo de 2021 en el caso

⁴² *People of the State of California v. Meta Platforms Inc.*, 4:23-cv 05448, Distrito Norte de California.

*Lemon v. Snap Inc.*⁴³, en la que sí se enjuicia la aplicabilidad de la sección 230 CDS a las características del diseño un producto de una red social, aunque estas estén relacionadas con el contenido de terceros.

54. Este caso se presentó a raíz de un violento accidente automovilístico a alta velocidad que mató a tres jóvenes en 2017. El accidente ocurrió cuando los chicos circulaban a más de 160 km/h por una carretera de Wisconsin, mientras documentaban su velocidad con el filtro de velocidad de Snapchat (*Speed Filter*)⁴⁴. Los padres de dos de los pasajeros fallecidos presentaron una demanda por negligencia en el diseño, alegando que la interacción entre el filtro de velocidad de Snapchat y su sistema de recompensas de «trofeos, rachas y reconocimientos sociales» representaba un defecto de diseño, que causó la muerte de sus hijos al fomentar la conducción peligrosa a alta velocidad.

55. El tribunal de distrito desestimó la demanda en aplicación de la sección 230 CDA, defendiendo la inmunidad de los operadores de servicios de la información cuando alojan, archivan, anuncian o no eliminan contenidos perjudiciales de terceros. Pero, el tribunal de circuito consideró que en este caso se estaba valorando si el daño se debía a algo distinto del contenido de un usuario externo. En este caso, el daño fue un accidente de coche supuestamente causado por la influencia del filtro de velocidad en los conductores adolescentes, y no por el mensaje de Snap que los chicos enviaron antes de su muerte. La cuestión para el tribunal, por lo tanto, era si los demandantes consideraban a Snap como el editor o emisor de los mensajes de los demandantes. El tribunal razonó que la reclamación de los demandantes no trataba a Snap como un editor de los mensajes de Snap de los chicos, sino más bien como el diseñador de un producto inseguro que publicaba el propio contenido del velocímetro de Snap. Snapchat podría ser considerada responsable de las violaciones del derecho de responsabilidad civil derivadas del contenido creado por su Filtro de velocidad. Además, el tribunal sostuvo que la demanda de los demandantes no alegaba una violación de un deber relacionado con el contenido o la publicación de los Snaps de las víctimas. Más bien, los demandantes alegaron una violación del deber de Snap de diseñar un producto razonablemente seguro, lo que el tribunal consideró independiente del papel de Snap en la supervisión y publicación de contenido de terceros.

56. Si se pasa el filtro de la CDA, el principal problema que las acciones enfrentan en estos casos, además, por supuesto, de la cuantificación del daño, es la prueba de la causalidad adecuada y suficiente. Por una parte, es necesario acreditar la existencia de adicción física a la plataforma (*general causation*). En este sentido, se están sucediendo los estudios e informes psicológicos, médicos, psiquiátricos, en los que trata de acreditarse su potencial adictivo. Otras pruebas indirectas o indiciarias del posible carácter adictivo de las redes sociales y de los efectos que para la salud mental, especialmente de adolescentes, pueden derivarse, del examen de los documentos internos de las propias plataformas⁴⁵.

57. Pero, además, aunque se pruebe la existencia de adicción es necesario acreditar el nexo causal entre esta y el daño mental o físico sufrido por el usuario (*specific causation*) lo que puede ser difícil, por ejemplo, en los casos de suicidio u otros daños multifactoriales, en los que pueden confluir varias causas.

58. Algún sector doctrinal ha defendido que podría facilitarse la prueba del nexo causal mediante la imposición de auditorías obligatorias para evaluar los riesgos asociados con los algoritmos utilizados por las plataformas de redes sociales⁴⁶. Estas auditorías buscan garantizar que los algoritmos no generen efectos negativos, como la propagación de desinformación, la discriminación o la creación de burbujas de filtro que puedan afectar a los usuarios. La idea es promover una mayor transparencia y responsabilidad en el uso de la tecnología, asegurando que las plataformas operen de manera ética y segura.

⁴³ *Lemmon v. Snap, Inc.*, 995 F.3d 1085.

⁴⁴ Snapchat es una red social que permite a sus usuarios hacer fotos o vídeos (conocidos coloquialmente como snaps) y compartirlos con otros usuarios de Snapchat. Para mantener a sus usuarios enganchados, Snapchat los recompensa con trofeos, rachas y reconocimientos sociales basados en los snaps que envían. La aplicación también permite a sus usuarios superponer un filtro sobre las fotos o vídeos que capturan a través de Snapchat en el momento en que realizan esa foto o vídeo.

⁴⁵ Vid. las referencias previas al caso *Facebook Files*.

⁴⁶ Vid. COSTELLO ET ALTRI: “Algorithms, Addiction and Adolescent Mental Health: *cit.*”, pp. 170 y ss.

C) Daños derivados del engaño: ocultamiento del carácter adictivo de la plataforma (*failure to warn claims*)

59. En tercer lugar, algunas de las pretensiones resarcitorias se apoyan en la supuesta deslealtad de la conducta de la plataforma, concretamente en su carácter engañoso por omisión, al no informar sobre la naturaleza adictiva de su producto y el riesgo de que este dé lugar a un uso compulsivo y otras lesiones mentales o físicas. Se defiende el interés del usuario o consumidor a no ser expuesto a un producto dañino sin su conocimiento y voluntad⁴⁷.

IV. ¿Y el futuro?... Armas existentes y soluciones regulatorias

1. *Nihil novum sub sole*

60. La idea de que la adicción y el comportamiento compulsivo crean un fallo de mercado que amenaza la salud pública y que los vendedores pueden explotar a los compradores haciéndoles engancharse a sus productos, no es nueva. La encontramos en la regulación de productos como el tabaco, el alcohol, las drogas, etc. El temor de que esta posibilidad de adicción pueda ser mayor en sujetos que aún no tienen su voluntad y capacidad plenamente desarrollada, tampoco es novedoso.

61. Existe un debate abierto sobre los posibles efectos perjudiciales de las redes sociales en la salud mental de sus usuarios, especialmente en el caso de los usuarios menores de edad por su especial vulnerabilidad. Es un debate en ciernes, que enfrenta a escépticos que niegan que el diseño adictivo de la plataforma sea una amenaza genuina para la salud pública o, al menos, niegan que hoy en día exista suficiente evidencia que establezca tal amenaza o que las intervenciones normativas impliquen beneficios suficientes como para justificar la intervención regulatoria en el mercado o, que sean de tan difícil implementación en la práctica que sea posible abogar por un *laissez-faire* y la ausencia de intervención. Las posturas más escépticas rechazan categóricamente la conveniencia de la intervención pública y defienden la capacidad correctora -si fuera necesaria- del libre mercado y de las decisiones individuales. De esta forma, las preocupaciones sobre las posibilidades y los riesgos de la adicción son reducidas a cuestiones de autonomía individual y autocontrol. Cualquier intervención para limitar o controlar el tiempo, los contenidos o el diseño de una plataforma es tachada de paternalista e intrusiva, al cuestionar las verdaderas decisiones y preferencias de los usuarios que libremente eligen y disfrutan de estos servicios adictivos⁴⁸.

62. Por otro lado, otras posturas defienden la necesidad de adoptar medidas específicas para salvaguardar la salud de los usuarios, de forma similar a las adoptadas con otros productos, servicios o actividades percibidos como adictivos y peligrosos. Se recela del funcionamiento corrector del libre mercado sobre la base de distintos argumentos. En primer lugar, se cuestiona la pretendida libertad y consciencia de los usuarios, denunciándose que, en muchos casos, son las propias empresas las que recurren a “la libertad de elección” de sus usuarios, utilizándola como una justificación ética del incremento de la vigilancia, la manipulación y la exposición al riesgo. Se recela de la existencia de una libertad real, porque esta es ahogada o eliminada por gran parte de los mecanismos y trucos utilizados por las empresas para generar el compromiso y captar la atención continuada del usuario, dado que estos interfieren con su habilidad para tomar decisiones racionales y suficientemente informadas.

63. Las empresas de redes sociales emplean incesantes fuentes de contenido basado en algoritmos para mantener a los usuarios jóvenes interesados en las plataformas, lo que genera miles de millones en ingresos anuales para la plataforma pagados por los anunciantes que dirigen anuncios a

⁴⁷ Como señala T. Wu: *The attention merchants*, cit., p. 7, “it is essential that we fully understand the deal”.

⁴⁸ Recogen las distintas posturas contrarias a una intervención, E. BIETTI: “The Data-Attention Imperative”, cit., pp. 37 y ss. N. RICHARDS/W. HARTZOG: “Against engament”, cit., pp. 1161

menores⁴⁹. Los sistemas utilizados para generar esa atención implican un riesgo de adicción, especialmente en usuarios más jóvenes. Pese a que pueda resultar tentador reducir estos trastornos y sus consecuencias dañinas a problemas de elección individual, delegando las soluciones a herramientas basadas en el mercado, con los incentivos económicos actuales, es muy poco creíble que las plataformas vayan a modificar o eliminar las prácticas que perjudican a los usuarios jóvenes de las redes sociales. Estas posturas defienden, por tanto, la necesidad de alterar esos incentivos y de corregir esos fallos de mercado.

64. Dos vías se están simultaneando para hacer frente a los peligros percibidos. Por una parte, como hemos examinado, la acción judicial, aún en sus estadios iniciales, exigiendo responsabilidad a las plataformas por los daños que hayan podido causar con su conducta.

65. Un segundo campo de actuación lo constituyen los intentos de establecer límites a estos modelos de negocios mediante la aprobación de normativa regulatoria. En este sentido, el campo inicial de actuación y, de nuevo, se pone de manifiesto la relación entre la economía de la atención y la economía de la vigilancia ha sido, principalmente, la aprobación de normativa sobre protección de datos. Como remedio para combatir los trastornos de atención se atiende a la parte de los datos de la ecuación, considerando al capitalismo de vigilancia como la causa raíz de estos trastornos. Hasta ahora, el enfoque de la política y la ley no ha sido prohibir las prácticas comerciales impulsadas por datos, sino aislar los datos como un objeto de regulación de otras partes de un negocio y otorgar a los individuos derechos de protección en relación a estos datos⁵⁰.

66. Progresivamente se han ido abriendo nuevos flancos de intervención normativa. Desde las medidas más drásticas, como la ley australiana que prohíbe el uso de redes sociales a menores de 16 años⁵¹ o la ley francesa que impide el acceso a redes a menores de 15 años sin autorización parental⁵² hasta soluciones normativas centradas en la esencia del modelo, los mecanismos de *engagement*, que buscan asegurar que las técnicas y mecanismos empleados no afecten a la seguridad y salud de sus usuarios, especialmente, de los usuarios más jóvenes⁵³.

67. Así, la propia Ley de Servicios Digitales o, en España, el recientísimo *Anteproyecto de Ley Orgánica para la protección de las personas menores de edad en los entornos digitales*, que intentan garantizar la seguridad y salud de los usuarios de productos digitales, implicando activamente a los prestadores del servicio en esa garantía, por ejemplo, mediante la exigencia de implantación de controles parentales o medidas de verificación de edad efectivas.

2. Reflexión final...

68. A los menores se les ha limitado el acceso a productos considerados perjudiciales o que pueden poner en riesgo su salud. Se abre ahora un nuevo terreno pantanoso como consecuencia de la digitalización y el auge de las redes sociales, con delicados conflictos de intereses.

⁴⁹ COSTELLO ET ALTRI: “Algorithms, Addiction and Adolescent Mental Health”, *cit.*, p. 137 reconocen como en 2022, las principales plataformas de redes sociales obtuvieron casi 11 mil millones de dólares en ingresos publicitarios de niños estadounidenses de cero a diecisiete años.

⁵⁰ La insuficiencia del control de datos para frenar los problemas de adicción ha sido cuestionada por la doctrina: N. RICHARDS/W. HARTZOG: “Against engagement”, *cit.*, pp. 1164 y ss.; E. BIETTI: “The Data-Attention Imperative”, *cit.*, pp. 52 y ss.

⁵¹ *Online Safety Amendment (Social Media Minimum Age) Act 2024*. Está prevista la entrada en vigor de las restricciones de edad en Diciembre de 2025.

⁵² Loi n° 2019-222 du 23 mars 2019 *relative à la protection des données personnelles des mineurs sur les réseaux sociaux*.

⁵³ N. RICHARDS/W. HARTZOG: “Against engagement”, *cit.*, pp. 1155 y ss., insisten en la necesidad de controlar legalmente los mecanismos de compromiso y sancionar los casos en los que estos lesionen los intereses de los consumidores, de las instituciones o de la sociedad o tengan un carácter engañoso. Vid. OECD Digital Economy Papers, núm. 363, 2024: *Towards Digital Safety by Design for Children*.

69. Ahora bien, es necesario tener en consideración que, a la hora de imputar responsabilidades, a mi juicio, no pesan igual los dos lados de la balanza, que enfrenta, por un lado, la protección de la seguridad y salud de los ciudadanos (especialmente de los menores por ser un colectivo especialmente vulnerable y ser susceptible de sufrir daños mayores) con la libertad de empresa, en su caso, libertad de expresión, y fomento de la innovación y desarrollo de nuevas actividades económicas digitales⁵⁴, sobre todo si se tiene en cuenta la finalidad exclusivamente lucrativa de una de las partes y la vieja máxima latina: *ubi emolumentus ibi onus*.

70. Ciertamente, las plataformas no están obligadas a promover el bien común con sus negocios y actividades empresariales. El juez MATEY, en su voto particular en el caso *Tawainna Anderson* lo refleja acertadamente: “La empresa puede decidir seleccionar el contenido que ofrece a los niños para enfatizar las virtudes más bajas, los gustos más viles. Puede negarse a utilizar un bien común para promover el bien común...”. Ahora bien, esto no implica que pueda permitírseles que causen daño para lucrarse. Recuperamos aquí otro viejo axioma latino, estructural a todo el sistema de responsabilidad civil, que es el principio del *alterum non laedere*.

⁵⁴ Este argumento, surgido en los inicios de la revolución digital, debería empezar a ser reconsiderado con seriedad, o, al menos a valorarse su alcance, no pudiendo usarse para justificar una suerte de inmunidad general de las plataformas frente a cualquier exigencia de responsabilidad.